

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-242154
(P2003-242154A)

(43) 公開日 平成15年8月29日 (2003.8.29)

(51) Int.Cl. ⁷	識別記号	F I	テ-リ-ト* (参考)
G 0 6 F 17/30	1 7 0	G 0 6 F 17/30	1 7 0 F 4 B 0 2 4
C 1 2 M 1/00		C 1 2 M 1/00	A 4 B 0 2 9
C 1 2 N 15/09	Z N A	G 0 6 F 17/60	1 2 6 E 5 B 0 7 5
G 0 6 F 17/60	1 2 6		1 2 6 Q
		C 1 2 N 15/00	Z N A A
審査請求 未請求 請求項の数46 O L (全 40 頁)			

(21) 出願番号 特願2002-40746 (P2002-40746)

(22) 出願日 平成14年2月18日 (2002.2.18)

(71) 出願人 500520628

セレスター・レキシコ・サイエンシズ株式
会社

千葉県千葉市美浜区中瀬1丁目3番地 幕
張テクノガーデンD17

(72) 発明者 上村 泰央

千葉県千葉市美浜区中瀬1丁目3番地 幕
張テクノガーデンD17 セレスター・レキ
シコ・サイエンシズ株式会社内

(74) 代理人 100089118

弁理士 酒井 宏明 (外1名)

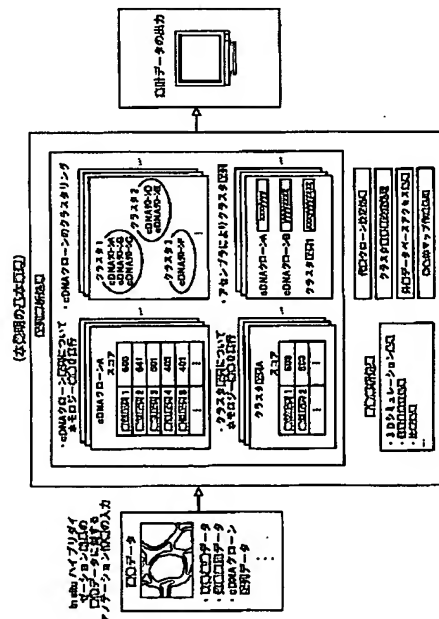
最終頁に続く

(54) 【発明の名称】 遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体

(57) 【要約】

【課題】 各種の遺伝子発現実験により取得した画像情報および遺伝子情報を統括的に管理し漏れなく知見を抽出する装置等を提供することを課題とする。

【解決手段】 図3に示すように、まず、本発明は、i n s i t uハイブリダイゼーション結果の画像データと対応する発現遺伝子 (c D N A) の塩基配列データを格納し、当該画像データに対するアノテーション情報 (採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる細胞などの部位に関する情報など) を利用者に入力させる。また、本発明は、これらのアノテーション情報について各種の画像解析技術を用いて画像データから自動的に認識する。



【特許請求の範囲】

【請求項1】 遺伝子の発現に関する画像データを入力する画像データ入力手段と、

上記発現した上記遺伝子の塩基配列を入力する塩基配列入力手段と、

上記塩基配列入力手段により入力された上記塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索手段と、

上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示手段と、

を備えたことを特徴とする遺伝子発現情報管理装置。

【請求項2】 上記ホモロジー検索手段は、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うこと、

を特徴とする請求項1に記載の遺伝子発現情報管理装置。

【請求項3】 遺伝子の発現に関する画像データを入力する画像データ入力手段と、

上記発現した上記遺伝子の塩基配列を入力する塩基配列入力手段と、

上記塩基配列入力手段により入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリング手段と、

上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示手段と、

を備えたことを特徴とする遺伝子発現情報管理装置。

【請求項4】 上記クラスタリング手段により同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定手段、

をさらに備え、

上記表示手段は、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩基配列とを表示する、ことを特徴とする請求項3に記載の遺伝子発現情報管理装置。

【請求項5】 上記クラスタリング手段は、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、

上記クラスタ配列決定手段は、上記クラスタの上記共通配列を上記クラスタ配列として決定する、

ことを特徴とする請求項3または4に記載の遺伝子発現情報管理装置。

【請求項6】 上記クラスタ配列決定手段により決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索手段、

をさらに備え、

上記表示手段は、上記クラスタ毎に上記クラスタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示する、

ことを特徴とする請求項4または5に記載の遺伝子発現情報管理装置。

【請求項7】 上記クラスタ配列ホモロジー検索手段は、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする請求項6に記載の遺伝子発現情報管理装置。

【請求項8】 上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納手段、

をさらに備え、

上記表示手段は、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示する、

ことを特徴とする請求項1から7のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項9】 上記画像データおよび/または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定手段、

をさらに備えたことを特徴とする請求項1から8のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項10】 上記発現量推定手段により推定された上記発現量に従って、上記画像データの表示順番をソートする発現量順ソート手段、

をさらに備えたことを特徴とする請求項9に記載の遺伝子発現情報管理装置。

【請求項11】 上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した

組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較手段と、

上記画像比較手段の比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する相違点抽出手段と、をさらに備えたことを特徴とする請求項1から10のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項12】 2つ以上の上記画像データから3次元画像を作成する3次元画像作成手段と、

上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーション手段と、

をさらに備えたことを特徴とする請求項1から11のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項13】 同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定手段、

をさらに備えたことを特徴とする請求項3から12のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項14】 上記クラスタに対して、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定手段、

をさらに備えたことを特徴とする請求項3から13のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項15】 上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定手段と、

上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成手段と、

上記染色体マップ作成手段にて作成した上記染色体マップを表示する染色体マップ表示手段と、

をさらに備えたことを特徴とする請求項1から14のいずれか一つに記載の遺伝子発現情報管理装置。

【請求項16】 遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、

上記塩基配列入力ステップにより入力された上記塩基配

列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索ステップと、

上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示ステップと、

を含むことを特徴とする遺伝子発現情報管理方法。

【請求項17】 上記ホモロジー検索ステップは、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うこと、

を特徴とする請求項16に記載の遺伝子発現情報管理方法。

【請求項18】 遺伝子の発現に関する画像データを入力する画像データ入力ステップと、

上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、

上記塩基配列入力ステップにより入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリングステップと、

上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示ステップと、

を含むことを特徴とする遺伝子発現情報管理方法。

【請求項19】 上記クラスタリングステップにより同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定ステップ、

をさらに含み、

上記表示ステップは、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩基配列とを表示する、

ことを特徴とする請求項18に記載の遺伝子発現情報管理方法。

【請求項20】 上記クラスタリングステップは、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、

上記クラスタ配列決定ステップは、上記クラスタの上記共通配列を上記クラスタ配列として決定する、

ことを特徴とする請求項18または19に記載の遺伝子発現情報管理方法。

【請求項21】 上記クラスタ配列決定ステップにより決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索ステップ、

をさらに含み、

上記表示ステップは、上記クラスタ毎に上記クラスタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示する、

ことを特徴とする請求項19または20に記載の遺伝子発現情報管理方法。

【請求項22】 上記クラスタ配列ホモロジー検索ステップは、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする請求項21に記載の遺伝子発現情報管理方法。

【請求項23】 上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納ステップ、

をさらに含む、

上記表示ステップは、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示する、

ことを特徴とする請求項16から22のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項24】 上記画像データおよび/または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定ステップ、

をさらに含むことを特徴とする請求項16から23のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項25】 上記発現量推定ステップにより推定された上記発現量に従って、上記画像データの表示順番をソートする発現量順ソートステップ、

をさらに含むことを特徴とする請求項24に記載の遺伝子発現情報管理方法。

【請求項26】 上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較ステップと、

上記画像比較ステップの比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する相違点抽出ステップと、

をさらに含むことを特徴とする請求項16から25のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項27】 2つ以上の上記画像データから3次元画像を作成する3次元画像作成ステップと、上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーションステップと、

をさらに含むことを特徴とする請求項16から26のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項28】 同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定ステップ、

をさらに含むことを特徴とする請求項18から27のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項29】 上記クラスタに対して、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定ステップ、

をさらに含むことを特徴とする請求項18から28のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項30】 上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定ステップと、上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成ステップと、

上記染色体マップ作成ステップにて作成した上記染色体マップを表示する染色体マップ表示ステップと、

をさらに含むことを特徴とする請求項16から29のいずれか一つに記載の遺伝子発現情報管理方法。

【請求項31】 遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、

上記塩基配列入力ステップにより入力された上記塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索ステップと、上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示ステップと、

を含む遺伝子発現情報管理方法をコンピュータに実行させることを特徴とするプログラム。

【請求項32】 上記ホモロジー検索ステップは、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うこと、

を特徴とする請求項31に記載のプログラム。

【請求項33】 遺伝子の発現に関する画像データを入力する画像データ入力ステップと、

上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、

上記塩基配列入力ステップにより入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリングステップと、

上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示ステップと、

を含む遺伝子発現情報管理方法をコンピュータに実行させることを特徴とするプログラム。

【請求項34】 上記クラスタリングステップにより同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定ステップ、

をさらに含み、

上記表示ステップは、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩基配列とを表示する、

ことを特徴とする請求項33に記載のプログラム。

【請求項35】 上記クラスタリングステップは、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、

上記クラスタ配列決定ステップは、上記クラスタの上記共通配列を上記クラスタ配列として決定する、

ことを特徴とする請求項33または34に記載のプログラム。

【請求項36】 上記クラスタ配列決定ステップにより決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索ステップ、

をさらに含み、

上記表示ステップは、上記クラスタ毎に上記クラスタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示する、

ことを特徴とする請求項34または35に記載のプログラム。

ラム。

【請求項37】 上記クラスタ配列ホモロジー検索ステップは、

(1) 既に同じあるいは別の生物で既知となっている遺伝子、

(2) 未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3) 未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、

(4) 染色体上での位置がわかっている遺伝子、

(5) 既に特許が取得されている遺伝子、

のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする請求項36に記載のプログラム。

【請求項38】 上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納ステップ、

をさらに含み、

上記表示ステップは、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示する、

ことを特徴とする請求項31から37のいずれか一つに記載のプログラム。

【請求項39】 上記画像データおよび/または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定ステップ、

をさらに含むことを特徴とする請求項31から38のいずれか一つに記載のプログラム。

【請求項40】 上記発現量推定ステップにより推定された上記発現量に従って、上記画像データの表示順序をソートする発現量順ソートステップ、

をさらに含むことを特徴とする請求項39に記載のプログラム。

【請求項41】 上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較ステップと、

上記画像比較ステップの比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する相違点抽出ステップと、

をさらに含むことを特徴とする請求項31から40のい

いずれか一つに記載のプログラム。

【請求項42】 2つ以上の上記画像データから3次元画像を作成する3次元画像作成ステップと、
上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーションステップと、

をさらに含むことを特徴とする請求項31から41のいずれか一つに記載のプログラム。

【請求項43】 同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定ステップ、

をさらに含むことを特徴とする請求項33から42のいずれか一つに記載のプログラム。

【請求項44】 上記クラスタに対して、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定ステップ、

をさらに含むことを特徴とする請求項33から43のいずれか一つに記載のプログラム。

【請求項45】 上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定ステップと、
上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成ステップと、

上記染色体マップ作成ステップにて作成した上記染色体マップを表示する染色体マップ表示ステップと、

をさらに含むことを特徴とする請求項31から44のいずれか一つに記載のプログラム。

【請求項46】 上記請求項31から45のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体に関し、特に、発現画像を管理する遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体に関する。

【0002】

【従来の技術】従来より、破碎していない細胞や組織に特定遺伝子のプローブを直接ハイブリダイズさせ、DN

AやmRNAなどの核酸あるいはタンパク質の細胞または組織内の分布や局在を光学顕微鏡や電子顕微鏡により調べる *in situ* ハイブリダイゼーションを用いて、遺伝子の発現場所などの同定を行う実験手法が広く行われている。

【0003】また、科学技術振興事業団土居バイオアシシメトリプロジェクトの小宮透らにより、臓器などの切片を用いた高速 *in situ* ハイブリダイゼーションの手法が開発されている（小宮透「96ウェルで行う切片の *in situ* ハイブリダイゼーション」細胞工学 18、405 1999年 など）。

【0004】この高速 *in situ* ハイブリダイゼーション法の概要について図1を参照して説明する。図1に示すように、まず、臓器などの組織からmRNAを精製し、NotIオリゴ(dT)を用いてmRNAからcDNAを合成してcDNAライブラリーを作製する（ステップSA-1）。そして、cDNAライブラリーの均一化を行い均一化ライブラリーを作製する（ステップSA-2）。本手法においては、cDNAの均一化ライブラリーを用いることにより、微量な遺伝子の発現についても同定することができる。

【0005】そして、ベクタへのライゲーションを行い、大腸菌へのトランスフォームを行う（ステップSA-3）。そして、トランスフォームされた大腸菌のコロニーを作製する（ステップSA-4）。そして、コロニーをランダムにピックアップし、ベクタの配列を使ってPCRにより内部のcDNAを直線化し増幅する。その際にはベクタ由来のRNAポリメラーゼのプロモーター配列が付加される。増幅されたcDNAをガラスパウダー法を用いて96ウェルプレートのフォーマットで精製し、これをマスターライブラリーとして保存する（ステップSA-5）。

【0006】そして、精製フラグメントを使って96ウェルプレートでハブテンであるDIG(digoxigenin)ラベルの転写反応を行う（ステップSA-6）。そして、出来上がったプローブをエタノール沈殿し、希釈して、96ウェルプレートの底にポリエステルワックスなどで貼り付けた臓器などの組織の切片に対して、*in situ* ハイブリダイゼーションを行う（ステップSA-7）。なお、ハイブリダイズ、酵素ラベルされた抗DIG抗体反応後の洗浄はELISA洗浄器を用いてシステムチックに行う。

【0007】そして、酵素抗体法による発色後、倒立顕微鏡で観察し、画像をCCDカメラで取り込む。ここで、図2は、取り込まれた画像の一例を示す図である。図2に示すように、発色（図中の黒色で示す部分）がある場所などの興味のある発現を示す像が得られた場合には、マスターライブラリーの対応するクローンの塩基配列を決定することで、遺伝子の発現像と塩基配列とを対応させて情報を得ることができる（ステップSA-

8)。

【0008】

【発明が解決しようとする課題】しかしながら、従来の手法によると、*in situ*ハイブリダイゼーションにより発現像を得て、プローブとして用いたcDNAの塩基配列の情報とのマッチングやその情報を用いた生物学的知見の発見は人手により行っているため、取得した情報を統括的に管理し漏れなく知見を抽出することが困難であるという問題点を有していた。以下、この問題点の内容について、一層具体的に説明する。

【0009】上述した高速*in situ*ハイブリダイゼーションでは、得られた大量の発現像の解析やその遺伝子配列のホモロジー検索などによる解析が人手により行われる。すなわち、実験者は、発現したcDNAクローンの塩基配列に基づいて、既存のcDNAデータベース（たとえば、ESTデータベースや完全長cDNAデータベースなど）に対してBLASTなどのホモロジー検索プログラムにより類似配列の検索を行い、発現した遺伝子(mRNA)の特定などを行う必要がある。また、cDNAライブラリーに含まれる複数のcDNAクローンにつき同一のmRNAに由来するものを抽出してクラスタリングを行ったり、同一のクラスタに含まれるcDNAクローンの発現画像の分類作業を行ったり、さらに、その画像から読取れる知見について抽出を行ったりする必要がある。

【0010】また、このような高速*in situ*ハイブリダイゼーション実験を一つの組織について行くと少なくとも数百から数千の画像が取得され、さらに近年のトランスクリプトーム解析やプロテオーム解析の進展に伴い、高速*in situ*ハイブリダイゼーションなどによる発現解析のための実験データは膨大になることが予想される。従って、人手による作業には自ずと限界が生ずるため、これらの作業につき統合的に管理することができるデータベースシステムの開発が望まれている。

【0011】また、*in situ*ハイブリダイゼーションに関する実験データを一元管理するデータベースシステムを導入してデータマイニング手法を駆使することにより、従来の人手による解析によっては発見することができなかった知見についてコンピュータシステム上で発見することができることが期待される。

【0012】なお、これまで説明した従来の技術および発明が解決しようとする課題は、上述した高速*in situ*ハイブリダイゼーションによる実験データを収集する場合に限られず、他の手法により遺伝子の発現情報を解析する全ての実験手法を用いる場合において、同様を考えることができる。

【0013】本発明は上記問題点に鑑みてなされたもので、各種の遺伝子発現実験により取得した画像情報および遺伝子情報を統括的に管理し漏れなく知見を抽出することのできる、遺伝子発現情報管理装置、遺伝子発現情

報管理方法、プログラム、および、記録媒体を提供することを目的としている。

【0014】

【課題を解決するための手段】このような目的を達成するため、請求項1に記載の遺伝子発現情報管理装置は、遺伝子の発現に関する画像データを入力する画像データ入力手段と、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力手段と、上記塩基配列入力手段により入力された上記塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索手段と、上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示手段とを備えたことを特徴とする。

【0015】この装置によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列（例えば、cDNAクローンの塩基配列など）を入力し、入力された塩基配列に対してホモロジー検索を行い類似塩基配列を抽出し、画像データと対応する塩基配列とその類似塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0016】すなわち、例えば*in situ*ハイブリダイゼーションなどの遺伝子発現実験において撮影された画像データに対応する、プローブとして用いたcDNAの塩基配列について、塩基配列データベース（ESTデータベースや完全長cDNAデータベースなど）に格納された既知の塩基配列とホモロジー検索（例えば、FastA、Blastなど）を行い、類似性の高い塩基配列を表示することにより、当該cDNAを特定し、当該cDNAと画像データとを併記して分かり易く表示することができるようになる。

【0017】また、類似塩基配列と一緒に当該塩基配列と類似塩基配列との間の類似度（例えば、ホモロジー検索のスコアなど）を表示してもよい。これにより、最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0018】さらに、類似配列に付属する情報として、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該cDNAの塩基配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と併記して表示してもよい。

【0019】また、請求項2に記載の装置は、請求項1に記載の遺伝子発現情報管理装置において、上記ホモロジー検索手段は、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の

遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0020】これはホモロジー検索の一例を一層具体的に示すものである。この装置によれば、ホモロジー検索は、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができるようになる。

【0021】すなわち、(1)から(5)の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0022】また、請求項3に記載の装置は、遺伝子の発現に関する画像データを入力する画像データ入力手段と、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力手段と、上記塩基配列入力手段により入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリング手段と、上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示手段とを備えたことを特徴とする。

【0023】この装置によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列(例えば、cDNAクローンの塩基配列など)を入力し、入力された塩基配列に対してクラスタリングを行い、塩基配列を特定のクラスタに分類し、クラスタ毎に画像データと対応する塩基配列とを表示するので、例えば同じmRNAに由来するcDNA(EST配列)を同じクラスタに分類すること等により、同一の性質を持つ塩基配列を集めて特定のクラスタに分類することができるようになる。

【0024】これにより、cDNAライブラリーからランダムに抽出したcDNAクローンの重複を排除すること等ができるようになり、さらに、実験結果のまとめ作業を効率化することができるようになる。

【0025】また、請求項4に記載の装置は、請求項3に記載の遺伝子発現情報管理装置において、上記クラスタリング手段により同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定手段をさらに備え、上記表示手段は、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩

基配列とを表示することを特徴とする。

【0026】この装置によれば、同一のクラスタに分類された塩基配列からクラスタ配列を決定し、クラスタ毎にクラスタ配列と画像データと対応する塩基配列とを表示するので、例えば、クラスタのメンバーの塩基配列を組み合わせで作成した塩基配列(例えば、完全長cDNAなど)をクラスタ配列として決定し、表示することができるようになる。

【0027】これにより、部分的なcDNA配列(EST配列)を用いた実験から、完全長cDNA(mRNA配列)などの遺伝子配列を容易に予想することができるようになる。

【0028】また、請求項5に記載の装置は、請求項3または4に記載の遺伝子発現情報管理装置において、上記クラスタリング手段は、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、上記クラスタ配列決定手段は、上記クラスタの上記共通配列を上記クラスタ配列として決定することを特徴とする。

【0029】この装置によれば、塩基配列を用いて共通配列のアセンブルを行い、同一の共通配列を構成した塩基配列を同一のクラスタに分類し、クラスタの共通配列をクラスタ配列として決定するので、アセンブル技術(短い断片配列から長い配列を作成する技術をいう。例えば、断片配列間の配列の重なりをマルチプルアライメントなどの手法を用いて探索し、重なりを持つ断片配列同士を合成することにより長い配列を生成する。)により部分的なcDNA配列からより完全長に近いcDNA配列を作成することができるようになる。

【0030】また、請求項6に記載の装置は、請求項4または5に記載の遺伝子発現情報管理装置において、上記クラスタ配列決定手段により決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索手段をさらに備え、上記表示手段は、上記クラスタ毎に上記クラスタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示することを特徴とする。

【0031】この装置によれば、決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出し、クラスタ毎にクラスタ配列とその類似配列と画像データと対応する塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0032】すなわち、プローブに用いたcDNAの塩基配列からアセンブルなどを実行することより合成したクラスタ配列について、塩基配列データベース(ESTデータベースや完全長cDNAデータベースなど)に格納された既知の塩基配列とホモロジー検索(例えば、FastA、Blastなど)を行い、類似性の高い塩基配列を表示することにより、発現した遺伝子を容易に特

定することができるようになる。

【0033】また、類似塩基配列と一緒に当該クラスタ配列と類似塩基配列との間の類似度（例えば、ホモロジースコアなど）を表示してもよい。これにより、クラスタ配列に最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0034】さらに、類似配列に付属する情報として、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該クラスタ配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と一緒に表示してもよい。

【0035】さらに、任意の遺伝子配列を入力することにより、入力された遺伝子について登録されたcDNAクローン配列とクラスタリングを行い、同一のクラスタに属するクラスタ配列に関する情報（画像データなど）を表示することが可能になる。

【0036】また、請求項7に記載の装置は、請求項6に記載の遺伝子発現情報管理装置において、上記クラスタ配列ホモロジー検索手段は、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0037】これはクラスタ配列に対するホモロジー検索の一例を一層具体的に示すものである。この装置によれば、ホモロジー検索は、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができるようになる。

【0038】すなわち、（1）から（5）の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0039】また、請求項8に記載の装置は、請求項1から7のいずれか一つに記載の遺伝子発現情報管理装置

において、上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納手段をさらに備え、上記表示手段は、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することを特徴とする。

【0040】これは画像データに対するアノテーション情報の一例を一層具体的に示すものである。この装置によれば、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納し、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することができるようになる。

【0041】また、請求項9に記載の装置は、請求項1から8のいずれか一つに記載の遺伝子発現情報管理装置において、上記画像データおよび／または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定手段をさらに備えたことを特徴とする。

【0042】この装置によれば、画像データおよび／または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定するので、発現パターン（均一に発現、偏って発現など）などを容易に特定することができるようになる。

【0043】ここで、画像データ中の蛍光色素などの発光量や発光面積を既知の画像解析手法などにより求めることにより、発現量を推定してもよい。また、画像データだけでなく塩基配列の情報も利用すれば、例えば、塩基配列にリピート配列が含まれている場合にはクロスハイブリダイズ（同じリピートを持つ当該mRNA以外のmRNAに対してもハイブリダイゼーション反応が起こる）の可能性が高いため推定された発現量に対する信用性は低い、などの評価を自動的にできるようになる。

【0044】また、例えば、同じ組織切片に対してプロンプトを発光させたものと発光させないものを撮影したり、別の種類の蛍光色素などで染色したものを撮影したりすることにより、画像解析処理を簡単に実行することができるようになる（その2枚の画像の差分をとることにより発現量を推定する等）。

【0045】また、請求項10に記載の装置は、請求項9に記載の遺伝子発現情報管理装置において、上記発現

量推定手段により推定された上記発現量に従って、上記画像データの表示順番をソートする発現量順ソート手段をさらに備えたことを特徴とする。

【0046】この装置によれば、推定された発現量に従って、画像データの表示順番をソートするので、効率的に実験結果を確認できるようになる。

【0047】また、請求項11に記載の装置は、請求項1から10のいずれか一つに記載の遺伝子発現情報管理装置において、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較手段と、上記画像比較手段の比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する相違点抽出手段とをさらに備えたことを特徴とする。

【0048】この装置によれば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出するので、画像間の差異を効率的に抽出することができるようになる。

【0049】これにより、画像認識あるいは手作業によってそれぞれの組織での発現パターンについてアノテーションを行ない、そのアノテーションを自動で比較し、相違がみられるものについて抽出して表示することができるようになる。

【0050】また、例えば、正常細胞と疾患細胞との比較や、発育段階または加齢段階の時系列での比較や、薬剤投与前と投与後の比較などを効率的に実行することができるようになる。

【0051】また、請求項12に記載の装置は、請求項1から11のいずれか一つに記載の遺伝子発現情報管理装置において、2つ以上の上記画像データから3次元画像を作成する3次元画像作成手段と、上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーション手段とをさらに備えたことを特徴とする。

【0052】この装置によれば、2つ以上の画像データから3次元画像を作成し、画像データの発現量から3次元画像における発現量をシミュレーションするので、一つの配列で臓器の全スライスをテストした場合、その画像をつなぐことにより臓器の3次元画像をシミュレーションし、各画像を画像解析して求めたmRNAの発現量を3次元的に補正し表示することができるようになる。

【0053】また、請求項13に記載の装置は、請求項3から12のいずれか一つに記載の遺伝子発現情報管理

装置において、同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定手段をさらに備えたことを特徴とする。

【0054】この装置によれば、同一のクラスタに属する塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定するので、例えば、同じmRNAに由来するクローン群のうち最もよい実験データがとれるクローンを一つ選んで代表クローンとして抽出することができるようになる。また、同じmRNA由来と思われるESTは同じクラスタに分類し、代表クローンのみを実験することによって、その後の実験回数を減らすことができる。つまり、クラスタリングによって、それまではEST数の実験が必要だったのに対し、代表クローン（クラスタ数存在する）についてののみ実験を行うことで足りるようになる。

【0055】ここで、同一mRNAに由来するcDNAクローンであっても切断箇所により配列が少しずつ異なる。in situハイブリダイゼーションを実施したときにそのcDNA配列によってきれいに発現シグナルがでるものとそうでないものがあるため、画像データをみてきれいにシグナルが出ているクローンを代表クローンとしてもよい。

【0056】また、cDNAクローンの塩基配列は、リピート配列を持たないものや、実験に適した配列長のもの代表クローンとしてもよい。

【0057】また、請求項14に記載の装置は、請求項3から13のいずれか一つに記載の遺伝子発現情報管理装置において、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定手段をさらに備えたことを特徴とする。

【0058】この装置によれば、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいてクラスタ重要度を決定

するので、これらの情報に基づいてクラスタの重要度を任意に決定し、利用者が注目するクラスタを容易に発見することができるようになる。

【0059】例えば、発現量と組織に関する情報に基づいて特定の発育段階または加齢段階の組織で高い発現を示すクローンについて重要度を高く決定してもよく、さらに、その塩基配列に対するホモロジー検索結果により既存の遺伝子配列DBにヒットしない(すなわち、既知の類似配列が存在しない)ものであれば、さらに重要度を高く決定してもよい。

【0060】また、請求項15に記載の装置は、請求項1から14のいずれか一つに記載の遺伝子発現情報管理装置において、上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定手段と、上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成手段と、上記染色体マップ作成ステップにて作成した上記染色体マップを表示する染色体マップ表示ステップとをさらに備えたことを特徴とする。

【0061】この装置によれば、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報(例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など)をマッピングすることにより染色体マップを作成することができるようになる。

【0062】また、当該染色体マップの遺伝子座に対応する部分(特定のマークなどを表示しても良い)を選択すると、その詳細な塩基配列に関する情報を表示するようになる。

【0063】また、本発明は遺伝子発現情報管理方法に関するものであり、請求項16に記載の遺伝子発現情報管理方法は、遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、上記塩基配列入力ステップにより入力された上記塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索ステップと、上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示ステップとを含むことを特徴とする。

【0064】この方法によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列(例えば、cDNAクローンの塩基配列など)を入力し、入力された塩基配列に対してホモロジー検索を行い類似塩基配列を抽出し、画像データと対応する塩基配列とその類似塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0065】すなわち、例えばin situハイブリダイゼーションなどの遺伝子発現実験において撮影され

た画像データに対応する、プローブとして用いたcDNAの塩基配列について、塩基配列データベース(ESTデータベースや完全長cDNAデータベースなど)に格納された既知の塩基配列とホモロジー検索(例えば、FastA、Blastなど)を行い、類似性の高い塩基配列を表示することにより、当該cDNAを特定し、当該cDNAと画像データとを併記して分かり易く表示することができるようになる。

【0066】また、類似塩基配列と一緒に当該塩基配列と類似塩基配列との間の類似度(例えば、ホモロジー検索のスコアなど)を表示してもよい。これにより、最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0067】さらに、類似配列に付随する情報として、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該cDNAの塩基配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と併記して表示してもよい。

【0068】また、請求項17に記載の方法は、請求項16に記載の遺伝子発現情報管理方法において、上記ホモロジー検索ステップは、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0069】これはホモロジー検索の一例を一層具体的に示すものである。この方法によれば、ホモロジー検索は、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができるようになる。

【0070】すなわち、(1)から(5)の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0071】また、請求項18に記載の方法は、遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、上記塩基配列入力ステップにより入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリングステップと、上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示ステップとを含むことを特徴とする。

【0072】この方法によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列（例えば、cDNAクローンの塩基配列など）を入力し、入力された塩基配列に対してクラスタリングを行い、塩基配列を特定のクラスタに分類し、クラスタ毎に画像データと対応する塩基配列とを表示するので、例えば同じmRNAに由来するcDNA（EST配列）を同じクラスタに分類すること等により、同一の性質を持つ塩基配列を集めて特定のクラスタに分類することができるようになる。

【0073】これにより、cDNAライブラリーからランダムに抽出したcDNAクローンの重複を排除すること等ができるようになり、さらに、実験結果のまとめ作業を効率化することができるようになる。

【0074】また、請求項19に記載の方法は、請求項18に記載の遺伝子発現情報管理方法において、上記クラスタリングステップにより同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定ステップをさらに含み、上記表示ステップは、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩基配列とを表示することを特徴とする。

【0075】この方法によれば、同一のクラスタに分類された塩基配列からクラスタ配列を決定し、クラスタ毎にクラスタ配列と画像データと対応する塩基配列とを表示するので、例えば、クラスタのメンバーの塩基配列を組み合わせで作成した塩基配列（例えば、完全長cDNAなど）をクラスタ配列として決定し、表示することができるようになる。

【0076】これにより、部分的なcDNA配列（EST配列）を用いた実験から、完全長cDNA（mRNA配列）などの遺伝子配列を容易に予想することができるようになる。

【0077】また、請求項20に記載の方法は、請求項18または19に記載の遺伝子発現情報管理方法において、上記クラスタリングステップは、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、上記クラスタ配列決定ステップは、上記クラスタの上記共通配列を上記クラスタ配列として決定することを特徴とする。

【0078】この方法によれば、塩基配列を用いて共通配列のアセンブルを行い、同一の共通配列を構成した塩基配列を同一のクラスタに分類し、クラスタの共通配列をクラスタ配列として決定するので、アセンブル技術（短い断片配列から長い配列を作成する技術をいう。例えば、断片配列間の配列の重なりをマルチプルアライメントなどの手法を用いて探索し、重なりを持つ断片配列同士を合成することにより長い配列を生成する。）により部分的なcDNA配列からより完全長に近いcDNA配列を作成することができるようになる。

【0079】また、請求項21に記載の方法は、請求項19または20に記載の遺伝子発現情報管理方法において、上記クラスタ配列決定ステップにより決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索ステップをさらに含み、上記表示ステップは、上記クラスタ毎に上記クラスタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示することを特徴とする。

【0080】この方法によれば、決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出し、クラスタ毎にクラスタ配列とその類似配列と画像データと対応する塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0081】すなわち、プローブに用いたcDNAの塩基配列からアセンブルなどを実行することより合成したクラスタ配列について、塩基配列データベース（ESTデータベースや完全長cDNAデータベースなど）に格納された既知の塩基配列とホモロジー検索（例えば、FastA、Blastなど）を行い、類似性の高い塩基配列を表示することにより、発現した遺伝子を容易に特定することができるようになる。

【0082】また、類似塩基配列と一緒に当該クラスタ配列と類似塩基配列との間の類似度（例えば、ホモロジースコアなど）を表示してもよい。これにより、クラスタ配列に最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0083】さらに、類似配列に付随する情報として、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該クラスタ配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と一緒に表示してもよい。

【0084】さらに、任意の遺伝子配列を入力することにより、入力された遺伝子について登録されたcDNAクローン配列とクラスタリングを行い、同一のクラスタ

に属するクラスタ配列に関する情報（画像データなど）を表示することが可能になる。

【0085】また、請求項22に記載の方法は、請求項21に記載の遺伝子発現情報管理方法において、上記クラスタ配列ホモロジー検索ステップは、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0086】これはクラスタ配列に対するホモロジー検索の一例を一層具体的に示すものである。この方法によれば、ホモロジー検索は、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、

（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができるようになる。

【0087】すなわち、（1）から（5）の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0088】また、請求項23に記載の方法は、請求項16から22のいずれか一つに記載の遺伝子発現情報管理方法において、上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納ステップをさらに含み、上記表示ステップは、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することを特徴とする。

【0089】これは画像データに対するアノテーション情報の一例を一層具体的に示すものである。この方法によれば、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも

一つを対応付けて格納し、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することができるようになる。

【0090】また、請求項24に記載の方法は、請求項16から23のいずれか一つに記載の遺伝子発現情報管理方法において、上記画像データおよび／または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定ステップをさらに含むことを特徴とする。

【0091】この方法によれば、画像データおよび／または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定するので、発現パターン（均一に発現、偏って発現など）などを容易に特定することができるようになる。

【0092】ここで、画像データ中の蛍光色素などの発光量や発光面積を既知の画像解析手法などにより求めることにより、発現量を推定してもよい。また、画像データだけでなく塩基配列の情報も利用すれば、例えば、塩基配列にリビート配列が含まれている場合にはクロスハイブリダイズ（同じリビートを持つ当該mRNA以外のmRNAに対してもハイブリダイゼーション反応が起こる）の可能性が高いため推定された発現量に対する信用性は低い、などの評価を自動的にできるようになる。

【0093】また、例えば、同じ組織切片に対してプロンプトを発光させたものと発光させないものを撮影したり、別の種類の蛍光色素などで染色したものを撮影したりすることにより、画像解析処理を簡単に実行することができるようになる（その2枚の画像の差分をとることにより発現量を推定する等）。

【0094】また、請求項25に記載の方法は、請求項24に記載の遺伝子発現情報管理方法において、上記発現量推定ステップにより推定された上記発現量に従って、上記画像データの表示順番をソートする発現量順ソートステップをさらに含むことを特徴とする。

【0095】この方法によれば、推定された発現量に従って、画像データの表示順番をソートするので、効率的に実験結果を確認できるようになる。

【0096】また、請求項26に記載の方法は、請求項16から25のいずれか一つに記載の遺伝子発現情報管理方法において、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較ステップと、上記画像比較ステップの比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する

相違点抽出ステップとをさらに含むことを特徴とする。

【0097】この方法によれば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出するので、画像間の差異を効率的に抽出することができるようになる。

【0098】これにより、画像認識あるいは手作業によってそれぞれの組織での発現パターンについてアノテーションを行ない、そのアノテーションを自動で比較し、相違がみられるものについて抽出して表示することができるようになる。

【0099】また、例えば、正常細胞と疾患細胞との比較や、発育段階または加齢段階の時系列での比較や、薬剤投与前と投与後の比較などを効率的に実行することができるようになる。

【0100】また、請求項27に記載の方法は、請求項16から26のいずれか一つに記載の遺伝子発現情報管理方法において、2つ以上の上記画像データから3次元画像を作成する3次元画像作成ステップと、上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーションステップとをさらに含むことを特徴とする。

【0101】この方法によれば、2つ以上の画像データから3次元画像を作成し、画像データの発現量から3次元画像における発現量をシミュレーションするので、一つの配列で臓器の全スライスをテストした場合、その画像をつなぐことにより臓器の3次元画像をシミュレーションし、各画像を画像解析して求めたmRNAの発現量を3次元的に補正し表示することができるようになる。

【0102】また、請求項28に記載の方法は、請求項18から27のいずれか一つに記載の遺伝子発現情報管理方法において、同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定ステップをさらに含むことを特徴とする。

【0103】この方法によれば、同一のクラスタに属する塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定するので、例えば、同じmRNAに由来するクローン群のうち最もよい実験データがとれるクローンを一

つ選んで代表クローンとして抽出することができるようになる。また、同じmRNA由来と思われるESTは同じクラスタに分類し、代表クローンのみを実験することによって、その後の実験回数を減らすことができる。つまり、クラスタリングによって、それまではEST数の実験が必要だったのに対し、代表クローン（クラスタ数存在する）についてのみ実験を行うことで足りるようになる。

【0104】ここで、同一mRNAに由来するcDNAクローンであっても切断箇所により配列が少しずつ異なる。in situハイブリダイゼーションを実施したときにそのcDNA配列によってきれいに発現シグナルがでるものとそうでないものがあるため、画像データをみてきれいにシグナルが出ているクローンを代表クローンとしてもよい。

【0105】また、cDNAクローンの塩基配列は、リビート配列を持たないものや、実験に適した配列長のもを代表クローンとしてもよい。

【0106】また、請求項29に記載の方法は、請求項18から28のいずれか一つに記載の遺伝子発現情報管理方法において、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定ステップをさらに含むことを特徴とする。

【0107】この方法によれば、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいてクラスタ重要度を決定するので、これらの情報に基づいてクラスタの重要度を任意に決定し、利用者が注目するクラスタを容易に発現することができるようになる。

【0108】例えば、発現量と組織に関する情報に基づいて特定の発育段階または加齢段階の組織で高い発現を示すクローンについて重要度を高く決定してもよく、さらに、その塩基配列に対するホモロジー検索結果により既存の遺伝子配列DBにヒットしない（すなわち、既知の類似配列が存在しない）ものであれば、さらに重要度を高く決定してもよい。

【0109】また、請求項30に記載の方法は、請求項16から29のいずれか一つに記載の遺伝子発現情報管理方法において、上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定ステップと、上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ

作成ステップと、上記染色体マップ作成ステップにて作成した上記染色体マップを表示する染色体マップ表示ステップとをさらに含むことを特徴とする。

【0110】この方法によれば、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報（例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など）をマッピングすることにより染色体マップを作成することができるようになる。

【0111】また、当該染色体マップの遺伝子座に対応する部分（特定のマークなどを表示しても良い）を選択すると、その詳細な塩基配列に関する情報を表示するようにしてもよい。

【0112】また、本発明はプログラムに関するものであり、請求項31に記載のプログラムは、遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、上記塩基配列入力ステップにより入力された上記塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索ステップと、上記画像データと対応する上記塩基配列とその上記類似塩基配列とを表示する表示ステップとを含むことを特徴とする。

【0113】このプログラムによれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列（例えば、cDNAクローンの塩基配列など）を入力し、入力された塩基配列に対してホモロジー検索を行い類似塩基配列を抽出し、画像データと対応する塩基配列とその類似塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0114】すなわち、例えば *in situ* ハイブリダイゼーションなどの遺伝子発現実験において撮影された画像データに対応する、プローブとして用いたcDNAの塩基配列について、塩基配列データベース（ESTデータベースや完全長cDNAデータベースなど）に格納された既知の塩基配列とホモロジー検索（例えば、FastA、Blastなど）を行い、類似性の高い塩基配列を表示することにより、当該cDNAを特定し、当該cDNAと画像データとを併記して分かり易く表示することができるようになる。

【0115】また、類似塩基配列と一緒に当該塩基配列と類似塩基配列との間の類似度（例えば、ホモロジー検索のスコアなど）を表示してもよい。これにより、最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0116】さらに、類似配列に付属する情報として、

遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該cDNAの塩基配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と併記して表示してもよい。

【0117】また、請求項32に記載のプログラムは、請求項31に記載のプログラムにおいて、上記ホモロジー検索ステップは、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0118】これはホモロジー検索の一例を一層具体的に示すものである。このプログラムによれば、ホモロジー検索は、（1）既に同じあるいは別の生物で既知となっている遺伝子、（2）未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、（3）未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、（4）染色体上での位置がわかっている遺伝子、（5）既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができるようになる。

【0119】すなわち、（1）から（5）の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0120】また、請求項33に記載のプログラムは、遺伝子の発現に関する画像データを入力する画像データ入力ステップと、上記発現した上記遺伝子の塩基配列を入力する塩基配列入力ステップと、上記塩基配列入力ステップにより入力された上記塩基配列に対してクラスタリングを行い、上記塩基配列を特定のクラスタに分類するクラスタリングステップと、上記クラスタ毎に上記画像データと対応する上記塩基配列とを表示する表示ステップとを含むことを特徴とする。

【0121】このプログラムによれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列（例えば、cDNAクローンの塩基配列など）を入力し、入力された塩基配列に対してクラスタリングを行い、塩基配列を特定のクラスタに分類し、クラスタ毎に画像データと対応する塩基配列とを表示するので、例え

ば同じ mRNA に由来する cDNA (EST 配列) を同じクラスタに分類すること等により、同一の性質を持つ塩基配列を集めて特定のクラスタに分類することができるようになる。

【0122】これにより、cDNA ライブラリーからランダムに抽出した cDNA クローンの重複を排除すること等ができるようになり、さらに、実験結果のまとめ作業を効率化することができるようになる。

【0123】また、請求項 34 に記載のプログラムは、請求項 33 に記載のプログラムにおいて、上記クラスタリングステップにより同一の上記クラスタに分類された上記塩基配列からクラスタ配列を決定するクラスタ配列決定ステップをさらに含み、上記表示ステップは、上記クラスタ毎に上記クラスタ配列と上記画像データと対応する上記塩基配列とを表示することを特徴とする。

【0124】このプログラムによれば、同一のクラスタに分類された塩基配列からクラスタ配列を決定し、クラスタ毎にクラスタ配列と画像データと対応する塩基配列とを表示するので、例えば、クラスタのメンバーの塩基配列を組み合わせで作成した塩基配列 (例えば、完全長 cDNA など) をクラスタ配列として決定し、表示することができるようになる。

【0125】これにより、部分的な cDNA 配列 (EST 配列) を用いた実験から、完全長 cDNA (mRNA 配列) などの遺伝子配列を容易に予想することができるようになる。

【0126】また、請求項 35 に記載のプログラムは、請求項 33 または 34 に記載のプログラムにおいて、上記クラスタリングステップは、上記塩基配列を用いて共通配列のアセンブルを行い、同一の上記共通配列を構成した上記塩基配列を同一の上記クラスタに分類し、上記クラスタ配列決定ステップは、上記クラスタの上記共通配列を上記クラスタ配列として決定することを特徴とする。

【0127】このプログラムによれば、塩基配列を用いて共通配列のアセンブルを行い、同一の共通配列を構成した塩基配列を同一のクラスタに分類し、クラスタの共通配列をクラスタ配列として決定するので、アセンブル技術 (短い断片配列から長い配列を作成する技術) をいう。例えば、断片配列間の配列の重なりをマルチプルアラインメントなどの手法を用いて探索し、重なりを持つ断片配列同士を合成することにより長い配列を生成する。) により部分的な cDNA 配列からより完全長に近い cDNA 配列を作成することができるようになる。

【0128】また、請求項 36 に記載のプログラムは、請求項 34 または 35 に記載のプログラムにおいて、上記クラスタ配列決定ステップにより決定された上記クラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索ステップをさらに含み、上記表示ステップは、上記クラスタ毎に上記クラ

スタ配列とその類似配列と上記画像データと対応する上記塩基配列とを表示することを特徴とする。

【0129】このプログラムによれば、決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出し、クラスタ毎にクラスタ配列とその類似配列と画像データと対応する塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができるようになる。

【0130】すなわち、プローブに用いた cDNA の塩基配列からアセンブルなどを実行することより合成したクラスタ配列について、塩基配列データベース (EST データベースや完全長 cDNA データベースなど) に格納された既知の塩基配列とホモロジー検索 (例えば、FastA、BLAST など) を行い、類似性の高い塩基配列を表示することにより、発現した遺伝子を容易に特定することができるようになる。

【0131】また、類似塩基配列と一緒に当該クラスタ配列と類似塩基配列との間の類似度 (例えば、ホモロジースコアなど) を表示してもよい。これにより、クラスタ配列に最も類似する塩基配列を表示したり、ホモロジースコアの高い順にソートして類似配列を表示したりすることができるようになる。

【0132】さらに、類似配列に付属する情報として、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子の GenBank データベースでの ID、当該遺伝子の産物タンパク質の GenBank データベースでの ID、当該クラスタ配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報のうち少なくとも一つを類似配列と一緒に表示してもよい。

【0133】さらに、任意の遺伝子配列を入力することにより、入力された遺伝子について登録された cDNA クローン配列とクラスタリングを行い、同一のクラスタに属するクラスタ配列に関する情報 (画像データなど) を表示することが可能になる。

【0134】また、請求項 37 に記載のプログラムは、請求項 36 に記載のプログラムにおいて、上記クラスタ配列ホモロジー検索ステップは、(1) 既に同じあるいは別の生物で既知となっている遺伝子、(2) 未知の遺伝子であるが同じ cDNA が既に取得されている遺伝子、(3) 未知の遺伝子であるが対応するゲノム DNA 断片が既に取得されている遺伝子、(4) 染色体上での位置がわかっている遺伝子、(5) 既に特許が取得されている遺伝子、のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うことを特徴とする。

【0135】これはクラスタ配列に対するホモロジー検索の一例を一層具体的に示すものである。このプログラムによれば、ホモロジー検索は、(1) 既に同じあるいは

は別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度を容易に特定することができるようになる。

【0136】すなわち、(1)から(5)の各カテゴリの塩基配列を格納した各塩基情報データベースを対象としてホモロジー検索を行うことができるようになるので、各カテゴリで最も類似度の高い塩基配列を個別に特定することができるようになる。

【0137】また、請求項38に記載のプログラムは、請求項31から37のいずれか一つに記載のプログラムにおいて、上記画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納するアノテーション情報格納ステップをさらに含み、上記表示ステップは、上記画像データに対して、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することを特徴とする。

【0138】これは画像データに対するアノテーション情報の一例を一層具体的に示すものである。このプログラムによれば、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納し、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することができるようになる。

【0139】また、請求項39に記載のプログラムは、請求項31から38のいずれか一つに記載のプログラムにおいて、上記画像データおよび/または上記塩基配列に基づいて、上記画像データ中の遺伝子の発現量を推定する発現量推定ステップをさらに含むことを特徴とする。

【0140】このプログラムによれば、画像データおよび/または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定するので、発現パターン(均一に発現、偏って発現など)などを容易に特定することができるようになる。

【0141】ここで、画像データ中の蛍光色素などの発光量や発光面積を既知の画像解析手法などにより求めることにより、発現量を推定してもよい。また、画像データだけでなく塩基配列の情報も利用すれば、例えば、塩基配列にリピート配列が含まれている場合にはクロスハイブリダイズ(同じリピートを持つ当該mRNA以外のmRNAに対してもハイブリダイゼーション反応が起こる)の可能性が高いため推定された発現量に対する信用性は低い、などの評価を自動的にできるようにする。

【0142】また、例えば、同じ組織切片に対してプローブを発光させたものと発光させないものを撮影したり、別の種類の蛍光色素などで染色したものを撮影したりすることにより、画像解析処理を簡単に実行することができるようになる(その2枚の画像の差分をとることにより発現量を推定する等)。

【0143】また、請求項40に記載のプログラムは、請求項39に記載のプログラムにおいて、上記発現量推定ステップにより推定された上記発現量に従って、上記画像データの表示順番をソートする発現量順ソートステップをさらに含むことを特徴とする。

【0144】このプログラムによれば、推定された発現量に従って、画像データの表示順番をソートするので、効率的に実験結果を確認できるようになる。

【0145】また、請求項41に記載のプログラムは、請求項31から40のいずれか一つに記載のプログラムにおいて、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の上記画像データを比較する画像比較ステップと、上記画像比較ステップの比較結果に基づいて、2つ以上の上記画像データの相違点を抽出する相違点抽出ステップとをさらに含むことを特徴とする。

【0146】このプログラムによれば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出するので、画像間の差異を効率的に抽出することができるようになる。

【0147】これにより、画像認識あるいは手作業によってそれぞれの組織での発現パターンについてアノテーションを行ない、そのアノテーションを自動で比較し、相違がみられるものについて抽出して表示することができるようになる。

【0148】また、例えば、正常細胞と疾患細胞との比較や、発育段階または加齢段階の時系列での比較や、薬

剤投与前と投与後の比較などを効率的に実行することができるようになる。

【0149】また、請求項42に記載のプログラムは、請求項31から41のいずれか一つに記載のプログラムにおいて、2つ以上の上記画像データから3次元画像を作成する3次元画像作成ステップと、上記画像データの上記発現量から上記3次元画像における発現量をシミュレーションする発現量シミュレーションステップとをさらに含むことを特徴とする。

【0150】このプログラムによれば、2つ以上の画像データから3次元画像を作成し、画像データの発現量から3次元画像における発現量をシミュレーションするので、一つの配列で臓器の全スライスをテストした場合、その画像をつなぐことにより臓器の3次元画像をシミュレーションし、各画像を画像解析して求めたmRNAの発現量を3次的に補正し表示することができるようになる。

【0151】また、請求項43に記載のプログラムは、請求項33から42のいずれか一つに記載のプログラムにおいて、同一の上記クラスタに属する上記塩基配列の中から、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定ステップをさらに含むことを特徴とする。

【0152】このプログラムによれば、同一のクラスタに属する塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定するので、例えば、同じmRNAに由来するクローン群のうち最もよい実験データがとれるクローンを一つ選んで代表クローンとして抽出することができるようになる。また、同じmRNA由来と思われるESTは同じクラスタに分類し、代表クローンのみを実験することによって、その後の実験回数を減らすことができる。つまり、クラスタリングによって、それまではEST数の実験が必要だったのに対し、代表クローン（クラスタ数存在する）についてのみ実験を行うことで足りるようになる。

【0153】ここで、同一mRNAに由来するcDNAクローンであっても切断箇所により配列が少しずつ異なる。in situハイブリダイゼーションを実施したときにそのcDNA配列によってきれいに発現シグナルがでるものとそうでないものがあるため、画像データをみてきれいにシグナルが出ているクローンを代表クローンとしてもよい。

【0154】また、cDNAクローンの塩基配列は、リピート配列を持たないものや、実験に適した配列長のものを代表クローンとしてもよい。

【0155】また、請求項44に記載のプログラムは、請求項33から43のいずれか一つに記載のプログラムにおいて、上記クラスタ配列のホモロジー検索結果、上記画像データ、上記塩基配列、上記発現量、上記採取した組織に関する情報、上記採取した組織の発育段階または加齢段階に関する情報、上記遺伝子の発現が見られるか否かに関する情報、および、上記発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定ステップをさらに含むことを特徴とする。

【0156】このプログラムによれば、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいてクラスタ重要度を決定するので、これらの情報に基づいてクラスタの重要度を任意に決定し、利用者が注目するクラスタを容易に見ることができるようになる。

【0157】例えば、発現量と組織に関する情報に基づいて特定の発育段階または加齢段階の組織で高い発現を示すクローンについて重要度を高く決定してもよく、さらに、その塩基配列に対するホモロジー検索結果により既存の遺伝子配列DBにヒットしない（すなわち、既知の類似配列が存在しない）ものであれば、さらに重要度を高く決定してもよい。

【0158】また、請求項45に記載のプログラムは、請求項31から44のいずれか一つに記載のプログラムにおいて、上記塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定ステップと、上記染色体の上記遺伝子座に上記塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成ステップと、上記染色体マップ作成ステップにて作成した上記染色体マップを表示する染色体マップ表示ステップとをさらに含むことを特徴とする。

【0159】このプログラムによれば、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報（例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など）をマッピングすることにより染色体マップを作成することができるようになる。

【0160】また、当該染色体マップの遺伝子座に対応する部分（特定のマークなどを表示しても良い）を選択すると、その詳細な塩基配列に関する情報を表示するようにしてもよい。

【0161】また、本発明は記録媒体に関するものであり、請求項46に記載の記録媒体は、上記請求項31から45のいずれか一つに記載されたプログラムを記録したことを特徴とする。

【0162】この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、請求項31から45のいずれか一つに記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

【0163】

【発明の実施の形態】以下に、本発明にかかる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。特に、以下の実施の形態においては、本発明を、高速 *in situ* ハイブリダイゼーションによる実験データを収集する場合に適用した例について説明するが、この場合に限られず、他の手法により遺伝子の発現情報を解析する全ての実験手法を用いる場合において、同様に適用することができる。

【0164】〔本発明の概要〕以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図3は本発明の基本原理を示す原理構成図である。

【0165】本発明は、概略的に、以下の基本的特徴を有する。すなわち、図3に示すように、まず、本発明は、*in situ* ハイブリダイゼーション結果の画像データと対応する発現遺伝子 (cDNA) の塩基配列データを格納し、当該画像データに対するアノテーション情報 (採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる細胞などの部位に関する情報など) を利用者に入力させる。また、本発明は、これらのアノテーション情報について既知の画像解析技術を用いて画像データから自動的に認識する。

【0166】そして、本発明は、入力されたcDNAの塩基配列に対して、塩基配列データベース (ESTデータベースや完全長cDNAデータベースなど) に格納された既知の塩基配列を対象としてホモロジー検索を行い、その類似塩基配列を抽出し、画像データと対応する塩基配列、その類似塩基配列、ホモロジスコア等を表示する。

【0167】また、本発明は、既に同じあるいは別の生物で既知となっている遺伝子、未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、染色体上での位置がわかっている遺伝子、既に特許が取得されている遺伝子のうち少なくとも一つの

遺伝子の塩基配列を対象として、cDNAクローン塩基の配列データとホモロジー検索を行う。

【0168】また、本発明は、例えば同じmRNAに由来するcDNA (EST配列) を同じクラスタに分類すること等により、同一の性質を持つ塩基配列を集めて特定のクラスタに分類する。また、同一のクラスタに分類された塩基配列からクラスタ配列を決定し、クラスタ毎にクラスタ配列と画像データと対応する塩基配列とを表示する。また、塩基配列を用いて共通配列のアセンブルを行い、同一の共通配列を構成した塩基配列を同一のクラスタに分類し、クラスタの共通配列をクラスタ配列として決定してもよい。

【0169】また、本発明は、決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出し、クラスタ毎にクラスタ配列とその類似配列と画像データと対応する塩基配列とを表示する。

【0170】また、本発明は、画像データおよび/または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定する。また、推定された発現量に従って、画像データの表示順番をソートしてもよい。

【0171】また、本発明は、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出する。

【0172】また、本発明は、2つ以上の画像データから3次元画像を作成し、画像データの発現量から3次元画像における発現量をシミュレーションする。

【0173】また、本発明は、同一のクラスタに属する塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する。

【0174】また、本発明は、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいてクラスタ重要度を決定する。

【0175】さらに、本発明は、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報 (例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に

関する情報など)をマッピングすることにより染色体マップを作成する。

【0176】[システム構成] まず、本システムの構成について説明する。図4は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。本システムは、概略的に、遺伝子発現情報管理装置100と、配列情報等に関する外部データベースやホモロジー検索等の外部プログラム等を提供する外部システム200とを、ネットワーク300を介して通信可能に接続して構成されている。

【0177】図4においてネットワーク300は、遺伝子発現情報管理装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネット等である。

【0178】図4において外部システム200は、ネットワーク300を介して、遺伝子発現情報管理装置100と相互に接続され、利用者に対してcDNAなどの配列情報等に関する外部データベースや、ホモロジー検索等の外部プログラムを実行するウェブサイトを提供する機能を有する。

【0179】ここで、外部システム200は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム200の各機能は、外部システム200のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

【0180】図4において遺伝子発現情報管理装置100は、概略的に、遺伝子発現情報管理装置100の全体を統括的に制御するCPU等の制御部102、通信回線等に接続されるルータ等の通信装置(図示せず)に接続される通信制御インターフェース部104、入力装置112や出力装置114に接続される入出力制御インターフェース部108、および、各種のデータベースやテーブルなどを格納する記憶部106を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。さらに、この遺伝子発現情報管理装置100は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク300に通信可能に接続されている。

【0181】記憶部106に格納される各種のデータベースやテーブル(cDNAクローンデータベース106a~核酸配列データベース106h)は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

【0182】これら記憶部106の各構成要素のうち、

cDNAクローンデータベース106aは、各cDNAクローンを一意に識別するcDNAクローン識別情報と、その塩基配列(EST配列)とを対応付けて格納する。

【0183】また、クラスタデータベース106bは、各クラスタを一意に識別するクラスタ識別情報と、クラスタを構成するcDNAクローンのcDNAクローン識別情報と、クラスタ配列を一意に識別するクラスタ配列識別情報と、代表クローンのcDNAクローン識別情報とを対応付けて格納する。

【0184】また、ホモロジー検索結果データベース106cは、cDNAクローン識別情報と、cDNAクローンの塩基配列に対して各種の塩基配列データベースに格納された塩基配列を対象としてホモロジー検索を行った検索結果とを対応付けて格納する。

【0185】また、クラスタ配列データベース106dは、クラスタ配列識別情報と、クラスタ配列の塩基配列とを対応付けて格納する。

【0186】また、クラスタ配列ホモロジー検索結果データベース106eは、クラスタ配列識別情報と、クラスタ配列の塩基配列に対して各種の塩基配列データベースに格納された塩基配列を対象としてホモロジー検索を行った検索結果とを対応付けて格納する。

【0187】また、ベクタ配列データベース106fは、cDNAクローン識別情報と、cDNAクローンを組み込んだベクタの塩基配列とを対応付けて格納する。

【0188】また、画像データベース106gは、画像データを一意に識別するための画像識別情報と、cDNAクローン識別情報と、アノテーション情報(例えば、遺伝子の発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など)とを対応付けて格納する。

【0189】また、核酸配列データベース106hは、EST配列や完全長cDNA配列などの塩基配列を格納したデータベースである。核酸配列データベース106hは、インターネットを経由してアクセスする外部の塩基配列データベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの配列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。ここで、核酸配列データベース106hは、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子に関する塩基配列を格納してもよい。

【0190】また、図4において、通信制御インターフェース部104は、遺伝子発現情報管理装置100とネットワーク300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部104は、他の端末と通信回線を介してデータを通信する機能を有する。

【0191】また、図4において、入出力制御インターフェース部108は、入力装置112や出力装置114の制御を行う。ここで、出力装置114としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置114をモニタとして記載する場合がある）。また、入力装置112としては、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

【0192】また、図4において、制御部102は、OS (Operating System) 等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部102は、機能概念的に、画像アノテーション情報入力部102a、cDNAクローン配列ホモロジー検索部102b、cDNAクローンクラスタリング部102c、クラスタ配列決定部102d、クラスタ配列ホモロジー検索部102e、3次元シミュレーション部102f、発現量推定部102g、画像比較部102h、代表クローン決定部102i、クラスタ重要度決定部102j、外部データベースアクセス部102k、染色体マップ作成部102m、および、表示画面作成部102nを備えて構成されている。

【0193】このうち、画像アノテーション情報入力部102aは、遺伝子の発現に関する画像データを入力する画像データ入力手段、および、発現した遺伝子の塩基配列を入力する塩基配列入力手段である。また、cDNAクローン配列ホモロジー検索部102bは、塩基配列入力手段により入力された塩基配列に対してホモロジー検索を行い類似塩基配列を抽出するホモロジー検索手段である。また、cDNAクローンクラスタリング部102cは、塩基配列入力手段により入力された塩基配列に対してクラスタリングを行い、塩基配列を特定のクラスタに分類するクラスタリング手段である。

【0194】また、クラスタ配列決定部102dは、クラスタリング手段により同一のクラスタに分類された塩基配列からクラスタ配列を決定するクラスタ配列決定手段である。また、クラスタ配列ホモロジー検索部102eは、クラスタ配列決定手段により決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出するクラスタ配列ホモロジー検索手段である。また、3次元シミュレーション部102fは、2つ以上の画像データから3次元画像を作成する3次元画像作成手段、お

よび、画像データの発現量から3次元画像における発現量をシミュレーションする発現量シミュレーション手段である。

【0195】また、発現量推定部102gは、画像データおよび/または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定する発現量推定手段、および、発現量推定手段により推定された発現量に従って、画像データの表示順番をソートする発現量順ソート手段である。また、画像比較部102hは、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較する画像比較手段、および、画像比較手段の比較結果に基づいて、2つ以上の画像データの相違点を抽出する相違点抽出手段である。

【0196】また、代表クローン決定部102iは、同一の上記クラスタに属する上記塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定する代表クローン決定手段である。

【0197】また、クラスタ重要度決定部102jは、クラスタに対して、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、クラスタ重要度を決定するクラスタ重要度決定手段である。また、外部データベースアクセス部102kは、ネットワーク300を介して外部システム200の外部データベースにアクセスする外部データベースアクセス手段である。

【0198】また、染色体マップ作成部102mは、塩基配列の存在する染色体上の遺伝子座を特定する遺伝子座特定手段、染色体の遺伝子座に塩基配列に関する情報をマッピングすることにより染色体マップを作成する染色体マップ作成手段、および、染色体マップ作成手段にて作成した染色体マップを表示する染色体マップ表示手段である。また、表示画面作成部102nは、画像データと対応する塩基配列とその類似塩基配列とを表示する表示手段、および、クラスタ毎に画像データと対応する塩基配列とを表示する表示手段である。なお、これら各部によって行なわれる処理の詳細については、後述する。

【0199】〔システムの処理〕次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図5～図13を参照して詳細に説明す

る。

【0200】〔画像アノテーション情報入力処理〕まず、画像アノテーション情報入力処理の詳細について図5を参照して説明する。図5は、本実施形態における本システムの画像アノテーション情報入力処理の一例を示すフローチャートである。

【0201】まず、遺伝子発現情報管理装置100は、画像アノテーション情報入力部102aの処理により、アノテーション情報入力画面を出力装置114に表示する(ステップSB-1)。

【0202】ここで、図13は、モニタに表示されるアノテーション情報入力画面の表示画面の一例を示す図である。この図に示すようにアノテーション情報入力画面は、例えば、画像識別情報の入力領域(図中の(1))、cDNAクローン識別情報の入力領域(図中の(2))、cDNAクローンの塩基配列決定が完了しているか否かを示す情報の入力領域(図中の(3))、切片を採取した組織名を入力領域(図中の(4))、切片を採取した生物の発育段階または加齢段階の入力領域(図中の(5))、どの細胞や組織などの部分領域に発現が見られるかを示す情報の入力領域(図中の(6))～(11))、切片上に遺伝子の発現が見られるか否かを示す情報の入力領域(図中の(12))～(15))、および、画像データの表示領域(図中の(16))を含んで構成されている。

【0203】について、ユーザがアノテーション情報入力画面を確認して各種の情報を入力装置112を介して入力すると(ステップSB-2)、遺伝子発現情報管理装置100は、画像アノテーション情報入力部102aの処理により、記憶部106の所定の記憶領域に入力情報を格納する(ステップSB-3)。これにて、画像アノテーション情報入力処理が終了する。

【0204】〔cDNAクローン配列ホモロジー検索処理〕次に、cDNAクローン配列ホモロジー検索処理の詳細について図6を参照して説明する。図6は、本実施形態における本システムのcDNAクローン配列ホモロジー検索処理の一例を示すフローチャートである。

【0205】まず、遺伝子発現情報管理装置100は、cDNAクローン配列ホモロジー検索部102bの処理により、cDNAクローンデータベース106aにアクセスしてcDNAクローン配列を取得する(ステップSC-1)。

【0206】について、cDNAクローン配列ホモロジー検索部102bは、核酸配列データベース106hにアクセスして、取得したcDNAクローン配列に対するホモロジー検索を実行する(ステップSC-2)。cDNAクローン配列ホモロジー検索部102bがアクセスする核酸配列データベース106hには、例えば、mam mal(哺乳動物)について既知の遺伝子配列を集めたデータベース、全生物について既知の遺伝子配列を集め

たデータベース、EST配列(cDNAクローンの断片配列)を集めたデータベース、ゲノム配列決定プロジェクトによって決定されつつあるゲノムDNA配列のドラフトを集めたデータベース、GSS配列(ゲノムDNAクローンの断片配列)を集めたデータベース、STS配列(ゲノム上でマッピングされた配列)を集めたデータベース、既に特許が取得されている遺伝子配列を集めたデータベースを含む。

【0207】について、cDNAクローン配列ホモロジー検索部102bは、ホモロジー検索結果(例えば、類似配列、ホモロジースコア、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該cDNAの塩基配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報などを含む)をホモロジー検索結果データベース106cに格納する(ステップSC-3)。これにて、cDNAクローン配列ホモロジー検索処理が終了する。

【0208】〔アセンブル処理〕次に、アセンブル処理の詳細について図7を参照して説明する。図7は、本実施形態における本システムのアセンブル処理の一例を示すフローチャートである。

【0209】まず、遺伝子発現情報管理装置100は、cDNAクローンクラスタリング部102cの処理により、cDNAクローンデータベース106aにアクセスして全てのcDNAクローンの塩基配列を取得し、既知のアセンブラ・ソフトウェアを用いて共通配列のアセンブルを行う(ステップSD-1)。

【0210】そして、クラスタ配列決定部102dは、複数のcDNAクローン配列からつながれた共通配列をクラスタ配列として決定し、クラスタ配列データベース106dの所定の記録領域に格納する(ステップSD-2)。

【0211】そして、cDNAクローンクラスタリング部102cは、同一のクラスタ配列を構成するcDNAクローンを同一クラスタに分類し、cDNAクローンデータベース106aにどのcDNAクローンがどのクラスタに含まれるか対応付けて記憶する(ステップSD-3)。これにて、アセンブル処理が終了する。

【0212】〔クラスタ配列ホモロジー検索処理〕次に、クラスタ配列ホモロジー検索処理の詳細について図8を参照して説明する。図8は、本実施形態における本システムのクラスタ配列ホモロジー検索処理の一例を示すフローチャートである。

【0213】まず、遺伝子発現情報管理装置100は、クラスタ配列ホモロジー検索部102eの処理により、クラスタ配列データベース106dにアクセスしてクラ

スタ配列を取得する(ステップSE-1)。

【0214】ついで、クラスタ配列ホモロジー検索部102eは、核酸配列データベース106hにアクセスして、取得したクラスタ配列に対するホモロジー検索を実行する(ステップSE-2)。クラスタ配列ホモロジー検索部102eがアクセスする核酸配列データベース106hには、例えば、mammal(哺乳動物)について既知の遺伝子配列を集めたデータベース、全生物について既知の遺伝子配列を集めたデータベース、EST配列(cDNAクローンの断片配列)を集めたデータベース、ゲノム配列決定プロジェクトによって決定されつつあるゲノムDNA配列のドラフトを集めたデータベース、GSS配列(ゲノムDNAクローンの断片配列)を集めたデータベース、STS配列(ゲノム上でマッピングされた配列)を集めたデータベース、既に特許が取得されている遺伝子配列を集めたデータベースを含む。

【0215】ついで、クラスタ配列ホモロジー検索部102eは、ホモロジー検索結果(例えば、類似配列、ホモロジスコア、遺伝子名、当該遺伝子の産物タンパク質名、当該遺伝子が取得されている生物種、当該遺伝子が取得された臓器名や組織名、当該遺伝子のGenBankデータベースでのID、当該遺伝子の産物タンパク質のGenBankデータベースでのID、当該クラスタ配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、または、当該遺伝子の存在の証拠に関する情報などを含む)をクラスタ配列ホモロジー検索結果データベース106eに格納する(ステップSE-3)。これにて、クラスタ配列ホモロジー検索処理が終了する。

【0216】[3次元シミュレーション処理]次に、3次元シミュレーション処理の詳細について図9を参照して説明する。図9は、本実施形態における本システムの3次元シミュレーション処理の一例を示すフローチャートである。

【0217】一つの配列で臓器の全スライスをテストした場合、遺伝子発現情報管理装置100は、3次元シミュレーション部102fの所により、2つ以上の画像データのセットを画像データベース106gから取得する(ステップSF-1)。

【0218】そして、3次元シミュレーション部102fは、画像データセットから既存の3D表示ソフトウェアなどを用いて3次元画像を作成する(ステップSF-2)。

【0219】そして、3次元シミュレーション部102fは、画像データの発現量から3次元画像における発現量をシミュレーションし、各画像を画像解析して求めた発現量を3次元的に補正し表示する(ステップSF-3)。これにて、3次元シミュレーション処理が終了する。

【0220】[発現量推定処理]次に、発現量推定処理

の詳細について図10を参照して説明する。図10は、本実施形態における本システムの発現量推定処理の一例を示すフローチャートである。

【0221】まず、遺伝子発現情報管理装置100は、発現量推定部102gの処理により、画像データベース106gおよびcDNAクローンデータベース106aにアクセスして、画像データおよび塩基配列を取得する(ステップSG-1)。

【0222】ついで、発現量推定部102gは、画像データおよび/または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定する。すなわち、発現量推定部102gは、画像データ中の蛍光色素などの発光量や発光面積を既知の画像解析手法などにより求めることにより(ステップSG-2)、発現量を推定する(ステップSG-3)。

【0223】また、画像データだけでなく塩基配列の情報も利用すれば、例えば、塩基配列にリピート配列が含まれている場合にはクロスハイブリダイズ(同じリピートを持つ当該mRNA以外のmRNAに対してもハイブリダイゼーション反応が起こる)の可能性が高いため推定された発現量に対する信用性は低い、などの評価を自動的にできるようになる。これにて、発現量推定処理が終了する。

【0224】[画像比較処理]次に、画像比較処理の詳細について図11を参照して説明する。図11は、本実施形態における本システムの画像比較処理の一例を示すフローチャートである。

【0225】まず、遺伝子発現情報管理装置100は、例えば、正常細胞と疾患細胞との比較や、発育段階または加齢段階の時系列での比較や、薬剤投与前と投与後の比較などを行う場合に、画像比較部102hの処理により、画像データベース106g等にアクセスして、比較される画像の組について、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報などを取得する(ステップSH-1)。

【0226】そして、画像比較部102hは、これらの情報に基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出する(ステップSH-2)。これにて、画像比較処理が終了する。

【0227】[染色体マップ作成処理]次に、染色体マップ作成処理の詳細について図12を参照して説明する。図12は、本実施形態における本システムの染色体マップ作成処理の一例を示すフローチャートである。

【0228】まず、遺伝子発現情報管理装置100は、染色体マップ作成部102mの処理により、cDNAクローンデータベース106aあるいはクラスタ配列データベース106d、および、核酸配列データベース10

6 hにアクセスすることにより、核酸配列データベース106 hに登録された遺伝子座の特定されている塩基配列情報をもとに、当該cDNAクローンあるいは当該クラスターの遺伝子座を特定する(ステップSJ-1)。

【0229】について、染色体マップ作成部102 mは、染色体の遺伝子座に塩基配列に関する情報(例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など)を染色体マップ上に配置(例えば、リンク情報を設定するなど)してマッピングする(ステップSJ-2)。

【0230】について、染色体マップ作成部102 mは、塩基配列に関する情報を追加して、当該染色体マップの遺伝子座に対応する部分(特定のマークなどを表示しても良い)を選択すると、その詳細な塩基配列に関する情報を表示するようにする(ステップSJ-3)。これにて、染色体マップ作成処理が終了する。

【0231】[実施例] 次に、このように構成された本発明の実施例について、以下に図14~図19を参照して詳細に説明する。

【0232】1. cDNAクローンごとのデータ閲覧
cDNAクローンごとのデータを閲覧する場合に表示される一覧レポート画面を図14に示す。図14に示すように、一覧レポート画面では、1行に1つのcDNAクローンについての情報が表示される。各列に表示される情報はそれぞれ以下の情報(1)~(13)を表わす(図14中の(1)~(13)を示す)。

【0233】(1) cDNAクローンIDを表示する。

【0234】(2) 当該cDNAクローンのGonadに対するin situ ハイブリダイゼーション画像が何枚存在するかを表示する。

【0235】(3) 当該cDNAクローンの生後1週のマウス精巣に対するin situ ハイブリダイゼーション画像が何枚存在するかを表示する。

【0236】(4) 当該cDNAクローンの生後2週のマウス精巣に対するin situ ハイブリダイゼーション画像が何枚存在するかを表示する。

【0237】(5) 当該cDNAクローンの生後3週のマウス精巣に対するin situ ハイブリダイゼーション画像が何枚存在するかを表示する。

【0238】(6) 当該cDNAクローンのAdultのマウス精巣に対するin situ ハイブリダイゼーション画像が何枚存在するかを表示する。

【0239】(7) 当該cDNAクローンのin situ ハイブリダイゼーション画像が全部で何枚存在するかを表示する。

【0240】(8) 当該cDNAクローンが所属クラスターの代表cDNAクローンである場合はYES、そうでない場合はNOを表示する。

【0241】(9) 当該cDNAクローンが属するクラスターのIDを表示する。

【0242】(10) 当該cDNAクローンのEST配列をnr(無冗長核酸配列データベース)に対してblast検索した結果のtop hit scoreを表示する。

【0243】(11) 当該cDNAクローンのEST配列をdbest(EST配列データベース)に対してblast検索した結果のtop hit scoreを表示する。

【0244】(12) 当該cDNAクローンのEST配列の長さ(塩基数)を表示する。

【0245】(13) 当該cDNAクローンのEST配列にベクタの配列が含まれている場合は「YES」、そうでない場合は「NO」を表示する。

【0246】次に詳細レポート画面について説明する。cDNAクローンごとのデータを閲覧する場合に表示される詳細レポート画面を図15に示す。図15に示すように、詳細レポート画面では、1つのcDNAクローンについての情報が1画面に表示される。図15中の各項目(1)~(26)で表示される情報を以下に示す。

【0247】(1) cDNAクローンIDを表示する。

【0248】(2) 当該cDNAクローンが属するクラスターのIDを表示する。

【0249】(3) 当該cDNAクローンが所属クラスターの代表cDNAクローンである場合は「YES」、そうでない場合は「NO」を表示する。

【0250】(4) 当該cDNAクローンのEST配列にベクタの配列が含まれている場合は「YES」、そうでない場合は「NO」を表示する。

【0251】(5)、(8)、(11)、(14)、(17) 当該cDNAクローンがin situ ハイブリダイゼーション実験によって、それぞれGonad、生後1週マウス精巣、生後2週マウス精巣、生後3週マウス精巣、Adultマウス精巣においてどの細胞に発現がみられたかを表示する。ここでG、C、R、E、L、Sはそれぞれ、「Gonia細胞」、「Spermatocyte細胞」、「Round spermatid細胞」、「Elongated spermatid細胞」、「Leydig細胞」、「Sertoli細胞」を表わす。

【0252】(6)、(9)、(12)、(15)、(18) 当該cDNAクローンのGonad、生後1週マウス精巣、生後2週マウス精巣、生後3週マウス精巣、Adultマウス精巣に対するin situ ハイブリダイゼーション画像がそれぞれ何枚存在するかを表示する。

【0253】(7)、(10)、(13)、(16)、

(19) 当該cDNAクローンのGonad、生後1週マウス精巢、生後2週マウス精巢、生後3週マウス精巢、Adultマウス精巢に対するin situ ハイブリダイゼーション画像をそれぞれ表示する。画像は画像IDと共に表示される。

【0254】(20) 当該cDNAクローンのEST配列をnrに対してblast検索した結果を表示する。

【0255】(21) 当該cDNAクローンのEST配列をnrに対してblast検索した結果のtop hit scoreを表示する。

【0256】(22) 当該cDNAクローンのEST配列をblast検索した対象nrデータのバージョン情報を表示する。

【0257】(23) 当該cDNAクローンのEST配列をdbestに対してblast検索した結果を表示する。

【0258】(24) 当該cDNAクローンのEST配列をdbestに対してblast検索した結果のtop hit scoreを表示する。

【0259】(25) 当該cDNAクローンのEST配列をblast検索した対象dbestデータのバージョン情報を表示する。

【0260】(26) 当該cDNAクローンのEST配列を表示する。

【0261】また、cDNAクローンごとのデータを閲覧する場合に表示される詳細レポート画面の別の一例を図16に示す。図16中の各項目(1)～(17)で表示される情報を以下に示す。

【0262】(1)～(5) 各発育段階または加齢段階での画像の右上にそれぞれ、「Y」、「N」、「?」、「Nd」というチェックボックスによって表示されている部分である、発現情報フィールドを表示する。これは画像から遺伝子の発現がみられるかどうかを入力するためのフィールドであり、「Y」は発現がみられること、「N」は発現がみられないこと、「?」はどちらとも判定できないこと、「Nd(No data)」は画像データが存在しないことを示す。

【0263】(6) 当該cDNAクローンの重要度を表示する。

【0264】(7) 既存のデータベースに掲載されたものであるか否かを表示する。

【0265】(8) 当該cDNAクローンのEST配列をホモロジー検索した結果のトップスコアをデータベース分類別に表示する。これにより当該クローンが未知のものであるか既知のものであるかが概観できる。ここで、「mammal」は哺乳動物について既知の遺伝子配列を集めたデータベース、「NT」は全生物について既知の遺伝子配列を集めたデータベース、「EST」はEST配列(cDNAクローンの断片配

列)を集めたデータベース、「HTG」はゲノム配列決定プロジェクトによって決定されつつあるゲノムDNA

配列のドラフトを集めたデータベース、「GSS」はGSS配列(ゲノムDNAクローンの断片配列)を集めたデータベース、「STS」はSTS配列(ゲノム上でマッピングされた配列)を集めたデータベース、「PAT」は既に特許が取得されている遺伝子配列を集めたデータベースを示す。この結果を見ることによって、当該cDNAクローンが、既に同じあるいは別の生物で既知となっている遺伝子(mammal、NT

)、未知の遺伝子であるが同じcDNAが既に取得されている遺伝子(EST)、未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子(HTG、GSS)、染色体上での位置がわかっている遺伝子(STS)、既に特許が取得されている遺伝子(PAT)であるかを確認することができる。

【0266】(9)～(17) 当該cDNAクローンが既知の遺伝子である場合、各フィールドにその既知遺伝子に関する情報が表示される。ここで、「Gene」は遺伝子名、「Product」は当該遺伝子の産物タンパク質名、「Organism」は当該遺伝子が取得されている生物種、「Tissue」は当該遺伝子が取得された臓器または組織、「Locus ID」は当該遺伝子のGenBankデータベースでのID、「Protein ID」は当該遺伝子の産物タンパク質のGenBankデータベースでのID、「Hit length、Hit Identity」は、当該cDNAの塩基配列が当該遺伝子配列にどのくらいの長さでどのくらい一致したか、「Evidence」は当該遺伝子の存在の証拠に関する情報(mRNA、DNA など;mRNA:mRNAの存在が確認されている、DNA:DNA配列から予測されたもの)を表示する。

【0267】2. クラスタごとのデータ閲覧
クラスタごとのデータを閲覧する場合に表示される一覧レポート画面を図17に示す。図17に示すように、一覧レポート画面では、1行に1つのクラスタについての情報が表示される。図17の各列はそれぞれ以下の情報(1)～(13)を表わす。

【0268】(1) クラスタIDを表示する。

【0269】(2) 当該クラスタに属するすべてのcDNAクローンのGonadに対するin situハイブリダイゼーション画像が合計何枚存在するかを表示する。

【0270】(3) 当該クラスタに属するすべてのcDNAクローンの生後1週のマウス精巢に対するin situハイブリダイゼーション画像が合計何枚存在するかを表示する。

【0271】(4) 当該クラスタに属するすべてのcDNAクローンの生後2週のマウス精巢に対するin situハイブリダイゼーション画像が合計何枚存在

するかを表示する。

【0272】(5) 当該クラスタに属するすべてのcDNAクローンの生後3週のマウス精巣に対する*in situ* ハイブリダイゼーション画像が合計何枚存在するかを表示する。

【0273】(6) 当該クラスタに属するすべてのcDNAクローンのAdultのマウス精巣に対する*in situ* ハイブリダイゼーション画像が合計何枚存在するかを表示する。

【0274】(7) 当該クラスタに属するすべてのcDNAクローンの*in situ* ハイブリダイゼーション画像が全部で何枚存在するかを表示する。

【0275】(8) 当該クラスタに属するcDNAクローンの数を表示する。

【0276】(9) 当該クラスタの代表cDNAクローンのIDを表示する。

【0277】(10) 当該クラスタの配列をnrに対してblast検索した結果のtop hit scoreを表示する。

【0278】(11) 当該クラスタの配列をdbestに対してblast検索した結果のtop hit scoreを表示する。

【0279】(12) 当該クラスタの重要度を表示する。重要度の高いものからAAA、AA、A、B、…のようにランク付けされている。

【0280】(13) 当該クラスタに属するいずれかのcDNAクローンのEST配列にベクタの配列が含まれている場合はYES、そうでない場合はNOを表示する。

【0281】クラスタごとのデータを閲覧する場合に表示される詳細レポート画面を図18に示す。図18に示すように、詳細レポート画面では、1つのクラスタについての情報が1画面に表示される。図中の各項目(1)～(29)に表示される情報は以下の通りである。

【0282】(1) クラスタIDを表示する。

【0283】(2) 当該クラスタの代表cDNAクローンのIDを表示する。

【0284】(3) 当該クラスタの重要度を表示する。重要度の高いものからAAA、AA、A、B、…のようにランク付けされている。

【0285】(4) 当該クラスタの配列および、所属cDNAクローンのEST配列について公開してもよいと思われるものにはYES、そうでないものにはNOを表示する。

【0286】(5) 当該クラスタに属するいずれかのcDNAクローンのEST配列にベクタの配列が含まれている場合はYES、そうでない場合はNOを表示する。

【0287】(6)、(9)、(12)、(15)、(18) 当該クラスタに属するcDNAクローンが

*in situ*ハイブリダイゼーション実験によって、それぞれGonad、生後1週マウス精巣、生後2週マウス精巣、生後3週マウス精巣、Adultマウス精巣においてどの細胞に発現がみられたかを表示する。ここでG、C、R、E、L、Sはそれぞれ、Gonia、Spermatocyte、Round spermatid、Elongated spermatid、Leydig cell、Sertoli cellを表わす。

【0288】(7)、(10)、(13)、(16)、(19) 当該クラスタに属するcDNAクローンのGonad、生後1週マウス精巣、生後2週マウス精巣、生後3週マウス精巣、Adultマウス精巣に対する*in situ* ハイブリダイゼーション画像がそれぞれ合計何枚存在するかを表示する。

【0289】(8)、(11)、(14)、(17)、(20) 当該クラスタに属するcDNAクローンのGonad、生後1週マウス精巣、生後2週マウス精巣、生後3週マウス精巣、Adultマウス精巣に対する*in situ* ハイブリダイゼーション画像をそれぞれ表示する。画像は画像IDと共に表示される。

【0290】(21) 当該クラスタに属するcDNAクローンの数を表示する。

【0291】(22) 当該クラスタに属するすべてのcDNAクローンのIDを表示する。各cDNAクローンはcDNAクローンIDとEST配列長(塩基数)が表示される。

【0292】(23) 当該クラスタの配列をnrに対してblast検索した結果を表示する。

【0293】(24) 当該クラスタの配列をnrに対してblast検索した結果のtop hit scoreを表示する。

【0294】(25) 当該クラスタの配列をblast検索した対象nrデータのバージョン情報を表示する。

【0295】(26) 当該クラスタの配列をdbestに対してblast検索した結果を表示する。

【0296】(27) 当該クラスタの配列をdbestに対してblast検索した結果のtop hit scoreを表示する。

【0297】(28) 当該クラスタの配列をblast検索した対象dbestデータのバージョン情報を表示する。

【0298】(29) 当該クラスタの配列を表示する。

【0299】3. 染色体マップの閲覧

染色体マップを閲覧する場合に表示される染色体マップ表示画面を図19に示す。図19に示すように、遺伝子座が特定されたcDNAクローン(またはクラスタ)が染色体マップ上に配置される。そして、そのcDNAク

ローンの発現量を時系列に一覧表示する。

【0300】また、各cDNAクローン（またはクラス名）をクリックなどすることにより選択すると、対応するcDNAクローンまたはクラスの詳細レポート画面が表示される。

【0301】[他の実施の形態] さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

【0302】例えば、遺伝子発現情報管理装置100がスタンドアローンの形態で処理を行う場合を一例に説明したが、遺伝子発現情報管理装置100とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

【0303】また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

【0304】また、遺伝子発現情報管理装置100に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。例えば、遺伝子発現情報管理装置100の各部または各装置が備える処理機能、特に制御部102にて行なわれる各処理機能については、その全部または任意の一部を、CPU（Central Processing Unit）および当該CPUにて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じて遺伝子発現情報管理装置100に機械的に読み取られる。

【0305】すなわち、ROMまたはHDなどの記憶部106などには、OS（Operating System）と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAM等にロードされることによって実行され、CPUと協働して制御部102を構成する。また、このコンピュータプログラムは、遺伝子発現情報管理装置100に対して任意のネットワーク300を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

【0306】また、本発明にかかるプログラムを、コン

ピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

【0307】また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS（Operating System）に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

【0308】記憶部106に格納される各種のデータベース等（cDNAクローンデータベース106a〜核酸配列データベース106h）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

【0309】また、遺伝子発現情報管理装置100は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

【0310】さらに、遺伝子発現情報管理装置100の分散・統合の具体的な形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部をCGI（Common Gateway Interface）を用いて実現してもよい。

【0311】また、ネットワーク300は、遺伝子発現情報管理装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VANや、パソコン通信網や、公衆電話網（アナログ／デ

デジタルの双方を含む)や、専用回線網(アナログ/デジタルの双方を含む)や、CATV網や、IMT2000方式、GSM方式またはPDC/PDC-P方式等の携帯回線交換網/携帯パケット交換網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含んでもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

【0312】

【発明の効果】以上詳細に説明したように、本発明によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列(例えば、cDNAクローンの塩基配列など)を入力し、入力された塩基配列に対してホモロジー検索を行い類似塩基配列を抽出し、画像データと対応する塩基配列とその類似塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0313】また、本発明によれば、ホモロジー検索は、(1)既に同じあるいは別の生物で既知となっている遺伝子、(2)未知の遺伝子であるが同じcDNAが既に取得されている遺伝子、(3)未知の遺伝子であるが対応するゲノムDNA断片が既に取得されている遺伝子、(4)染色体上での位置がわかっている遺伝子、(5)既に特許が取得されている遺伝子のうち少なくとも一つの遺伝子の塩基配列を対象としてホモロジー検索を行うので、画像データの持つ生物学的な重要度などを容易に特定することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0314】また、本発明によれば、遺伝子の発現に関する画像データを入力し、発現した遺伝子の塩基配列(例えば、cDNAクローンの塩基配列など)を入力し、入力された塩基配列に対してクラスタリングを行い、塩基配列を特定のクラスタに分類し、クラスタ毎に画像データと対応する塩基配列とを表示するので、例えば同じmRNAに由来するcDNA(EST配列)を同じクラスタに分類すること等により、同一の性質を持つ塩基配列を集めて特定のクラスタに分類することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0315】また、本発明によれば、同一のクラスタに分類された塩基配列からクラスタ配列を決定し、クラスタ毎にクラスタ配列と画像データと対応する塩基配列とを表示するので、例えば、クラスタのメンバーの塩基配列を組み合わせて作成した塩基配列(例えば、完全長cDNAなど)をクラスタ配列として決定し、表示するこ

とができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0316】また、本発明によれば、塩基配列を用いて共通配列のアセンブルを行い、同一の共通配列を構成した塩基配列を同一のクラスタに分類し、クラスタの共通配列をクラスタ配列として決定するので、アセンブル技術(短い断片配列から長い配列を作成する技術をいう。例えば、断片配列間の配列の重なりをマルチプルアライメントなどの手法を用いて探索し、重なりを持つ断片配列同士を合成することにより長い配列を生成する。)により部分的なcDNA配列からより完全長に近いcDNA配列を作成することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0317】また、本発明によれば、決定されたクラスタ配列に対してホモロジー検索を行い類似塩基配列を抽出し、クラスタ毎にクラスタ配列とその類似配列と画像データと対応する塩基配列とを表示するので、画像データにおいて発現した遺伝子を容易に特定することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0318】また、本発明によれば、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて格納し、画像データに対して、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つを対応付けて表示することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0319】また、本発明によれば、画像データおよび/または塩基配列に基づいて、画像データ中の遺伝子の発現量を推定するので、発現パターン(均一に発現、偏って発現など)などを容易に特定することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0320】また、本発明によれば、推定された発現量に従って、画像データの表示順番をソートするので、効率的に実験結果を確認できる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0321】また、本発明によれば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見ら

れる部位に関する情報のうち少なくとも一つに基づいて、2つ以上の画像データを比較し、比較結果に基づいて、2つ以上の画像データの相違点を抽出するので、画像間の差異を効率的に抽出することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0322】また、本発明によれば、2つ以上の画像データから3次元画像を作成し、画像データの発現量から3次元画像における発現量をシミュレーションするので、一つの配列で臓器の全スライスをテストした場合、その画像をつなぐことにより臓器の3次元画像をシミュレーションし、各画像を画像解析して求めたmRNAの発現量を3次元的に補正し表示することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0323】また、本発明によれば、同一のクラスタに属する塩基配列の中から、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいて、代表クローンを決定するので、例えば、同じmRNAに由来するクローン群のうち最もよい実験データがとれるクローンを一つ選んで代表クローンとして抽出することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0324】また、本発明によれば、クラスタ配列のホモロジー検索結果、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報のうち少なくとも一つに基づいてクラスタ重要度を決定するので、これらの情報に基づいてクラスタの重要度を任意に決定し、利用者が注目するクラスタを容易に発見することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【0325】さらに、本発明によれば、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報（例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など）をマッピングすることにより染色体マップを作成することができる遺伝子発現情報管理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【図面の簡単な説明】

【図1】高速in situハイブリダイゼーション法の概要を示す図である。

【図2】取り込まれた画像の一例を示す図である。

【図3】本発明の基本原理を示す原理構成図である。

【図4】本発明が適用される本システムの構成の一例を示すブロック図である。

【図5】本実施形態における本システムの画像アノテーション情報入力処理の一例を示すフローチャートである。

【図6】本実施形態における本システムのcDNAクローン配列ホモロジー検索処理の一例を示すフローチャートである。

【図7】本実施形態における本システムのアセンブル処理の一例を示すフローチャートである。

【図8】本実施形態における本システムのクラスタ配列ホモロジー検索処理の一例を示すフローチャートである。

【図9】本実施形態における本システムの3次元シミュレーション処理の一例を示すフローチャートである。

【図10】本実施形態における本システムの発現量推定処理の一例を示すフローチャートである。

【図11】本実施形態における本システムの画像比較処理の一例を示すフローチャートである。

【図12】本実施形態における本システムの染色体マップ作成処理の一例を示すフローチャートである。

【図13】モニタに表示されるアノテーション情報入力画面の表示画面の一例を示す図である。

【図14】cDNAクローンごとのデータを閲覧する場合に表示される一覧レポート画面の一例を示す図である。

【図15】cDNAクローンごとのデータを閲覧する場合に表示される詳細レポート画面の一例を示す図である。

【図16】cDNAクローンごとのデータを閲覧する場合に表示される詳細レポート画面の別の一例を示す図である。

【図17】クラスタごとのデータを閲覧する場合に表示される一覧レポート画面の一例を示す図である。

【図18】クラスタごとのデータを閲覧する場合に表示される詳細レポート画面の一例を示す図である。

【図19】染色体マップを閲覧する場合に表示される染色体マップ表示画面の一例を示す図である。

【符号の説明】

- 100 遺伝子発現情報管理装置
- 102 制御部
- 102a 画像アノテーション情報入力部
- 102b cDNAクローン配列ホモロジー検索部
- 102c cDNAクローンクラスタリング部
- 102d クラスタ配列決定部
- 102e クラスタ配列ホモロジー検索部
- 102f 3次元シミュレーション部
- 102g 発現量推定部

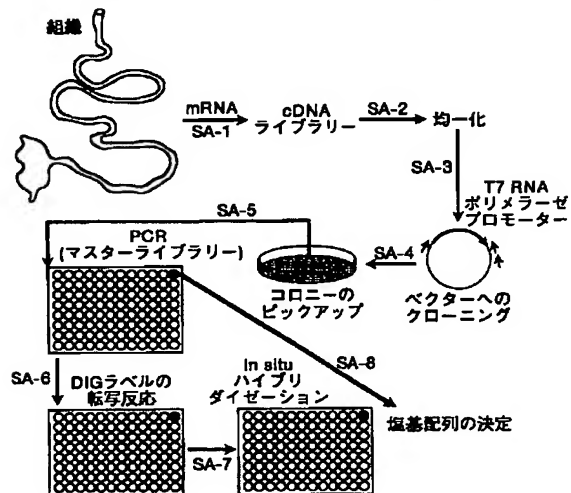
102h 画像比較部
 102i 代表クローン決定部
 102j クラスタ重要度決定部
 102k 外部データベースアクセス部
 102m 染色体マップ作成部
 102n 表示画面作成部
 104 通信制御インターフェース部
 106 記憶部
 106a cDNAクローンデータベース
 106b クラスタデータベース
 106c ホモロジー検索結果データベース

106d クラスタ配列データベース
 106e クラスタ配列ホモロジー検索結果データベース
 106f ベクタ配列データベース
 106g 画像データベース
 106h 核酸配列データベース
 108 入出力制御インターフェース部
 112 入力装置
 114 出力装置
 200 外部システム
 300 ネットワーク

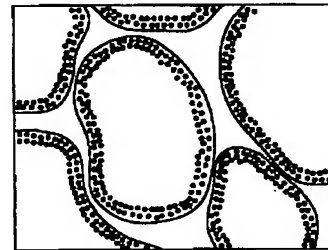
【図1】

【図2】

(高速 In situ ハイブリダイゼーションの手順)

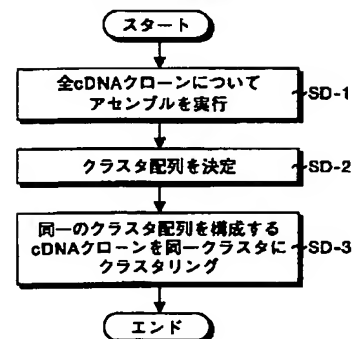


(In situ ハイブリダイゼーション結果の画像データの一例)



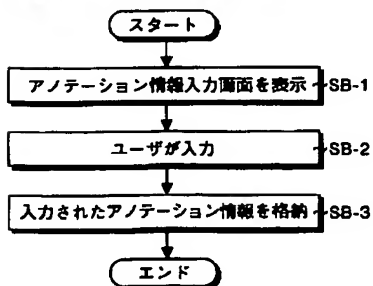
【図7】

(アセンブル処理)



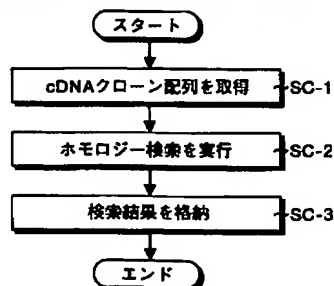
【図5】

(画像アノテーション情報入力処理)

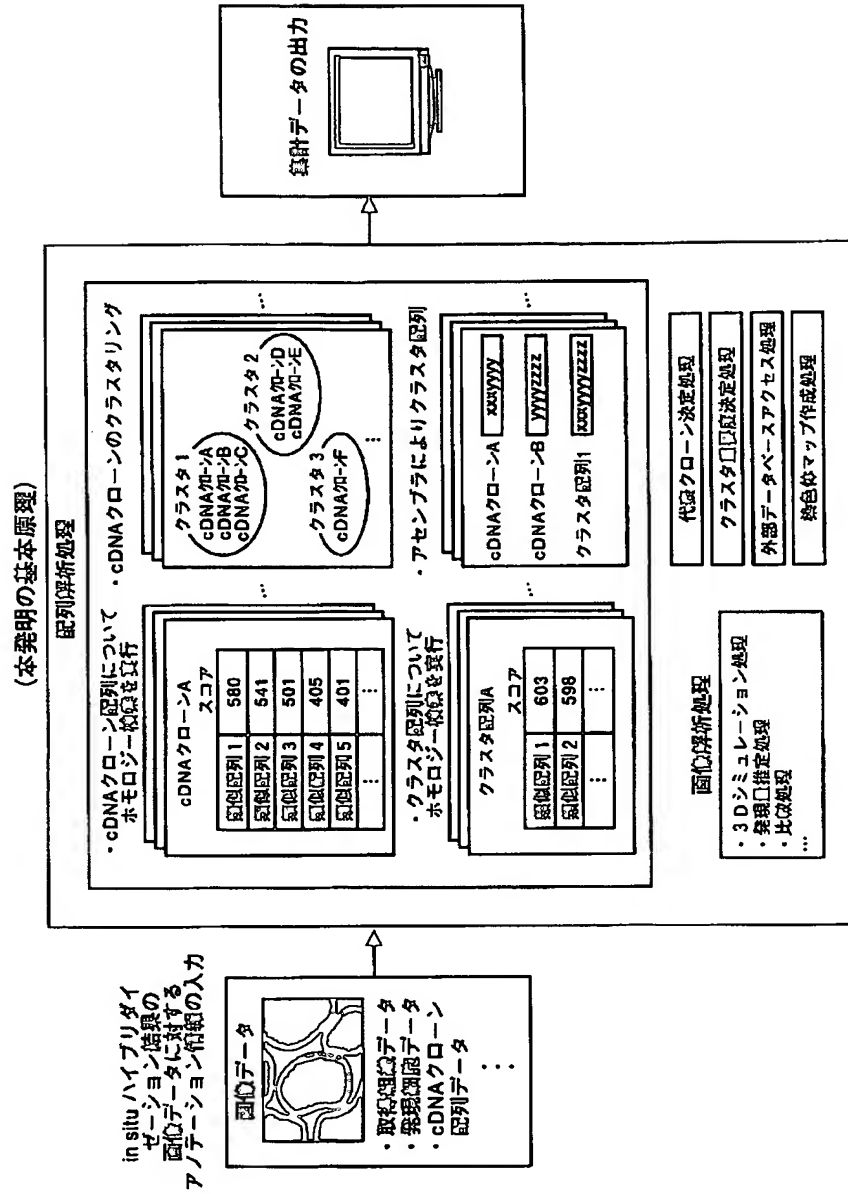


【図6】

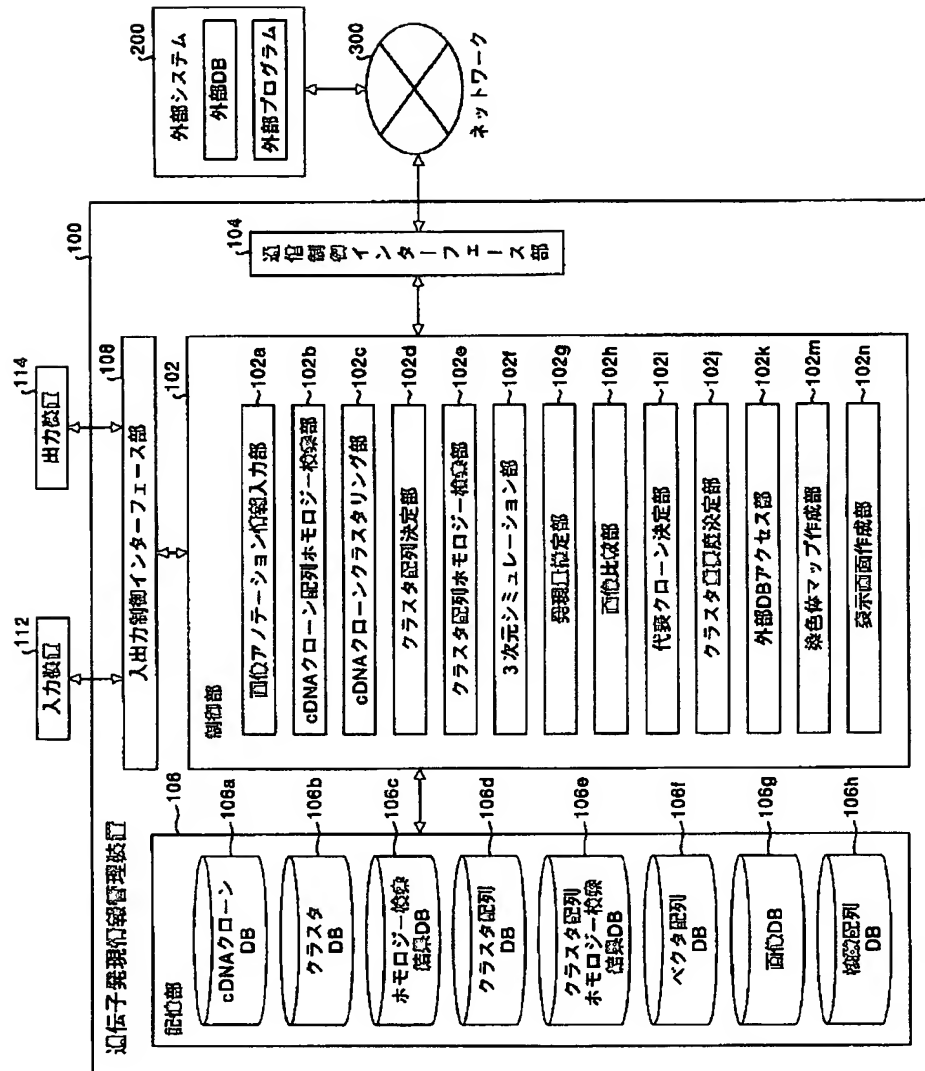
(cDNAクローン配列ホモロジー検索処理)



【図3】

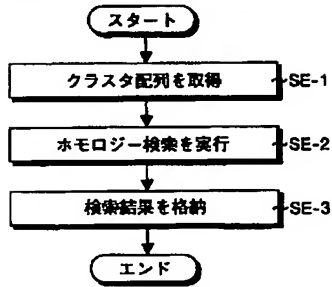


【図4】



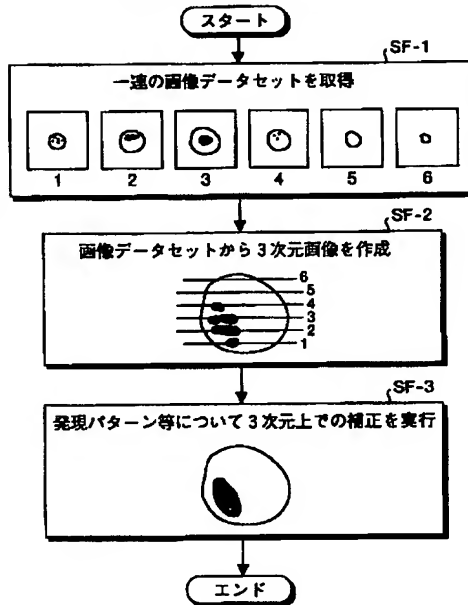
【図8】

(クラスタ配列ホモロジー検索処理)



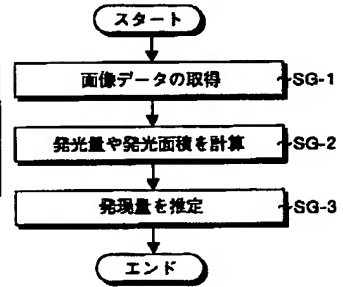
【図9】

(3次元シミュレーション処理)



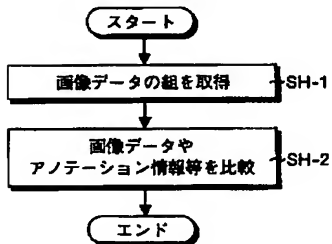
【図10】

(発現量推定処理)



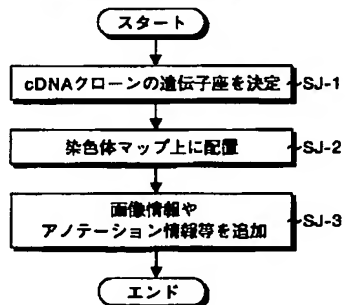
【図11】

(画像比較処理)



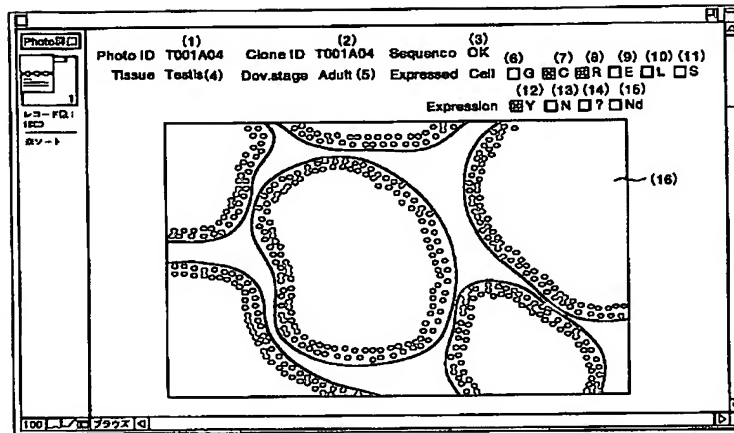
【図12】

(染色体マップ作成処理)



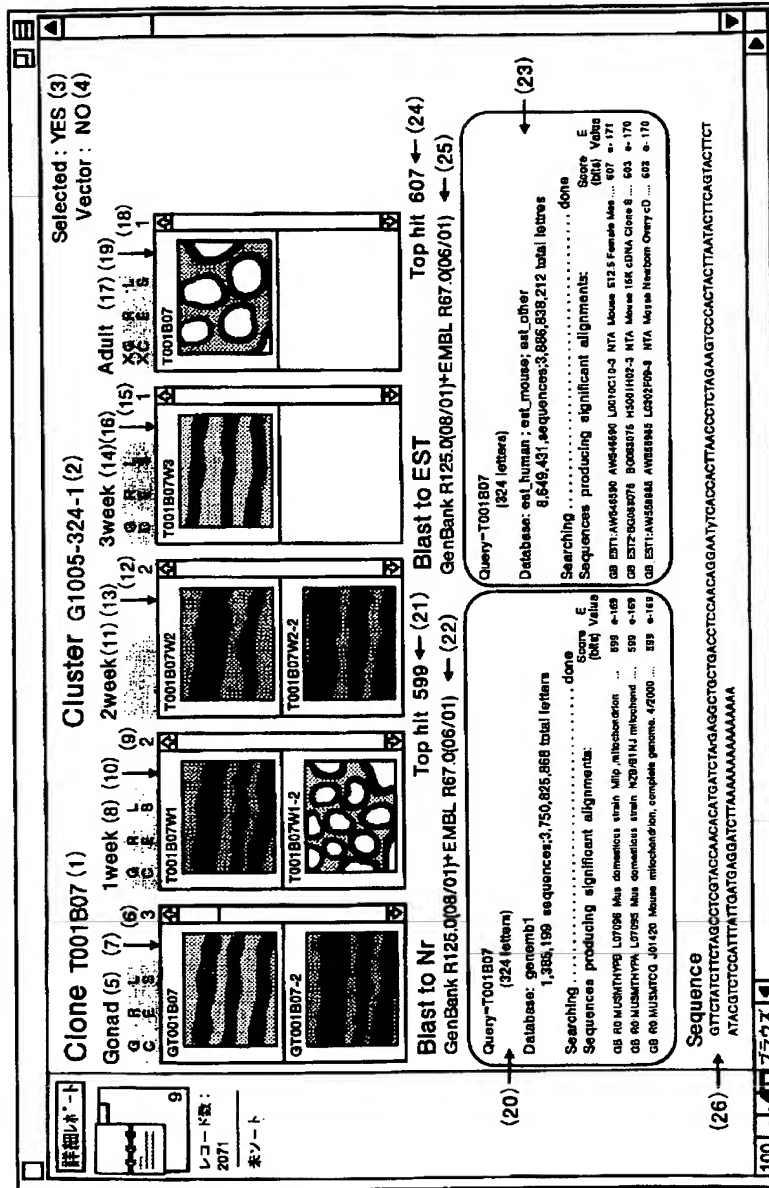
(84) 103-242154 (P2003-ch0味穀

【図13】



【図14】

clone View												
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Clone	Go	1w	2w	3w	Ad	Total	Sel	Cluster	Top hit (nr)	Top hit (EST)	Length	Vector
T001A04	0	2	2	2	1	7	YES	G0077-434-2	674	781	410	NO
T001A08	0	1	2	2	1	6	YES	G0012-526-16	967	969	508	NO
T001A09	2	2	2	1	0	7	NO	G0176-339-5			340	NO
T001A10	2	2	2	2	0	8	NO	G0008-549-21	635	635	436	NO
T001A12	0	1	2	2	1	6	YES	G0223-448-2	373	557	205	NO
T001B02	0	1	1	2	1	5	YES	G0062-420-3	601	601	419	NO
T001B04	3	2	2	2	1	10	YES	G0273-450-4	846	846	450	NO
T001B05	5	2	2	2	1	12	YES	G1006-258-1	480	480	258	NO
T001B07	3	2	2	1	1	9	YES	G1005-324-1	609	607	324	NO
T001B09	0	1	1	1	1	4	YES	G0145-353-10		652	340	NO
T001B10	0	1	1	2	1	5	YES	G0039-451-6	737	737	421	NO
T001C05	6	2	2	2	1	12	YES	G0090-541-8	741	741	541	NO



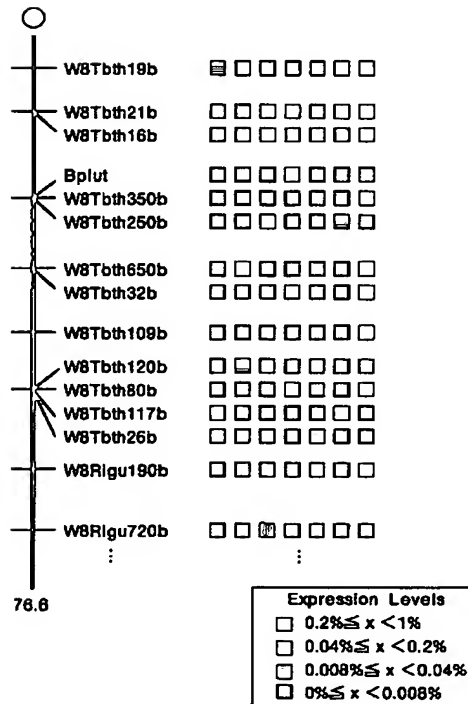
[illegible]

(37) 103-242154 (P2003-ch升級)

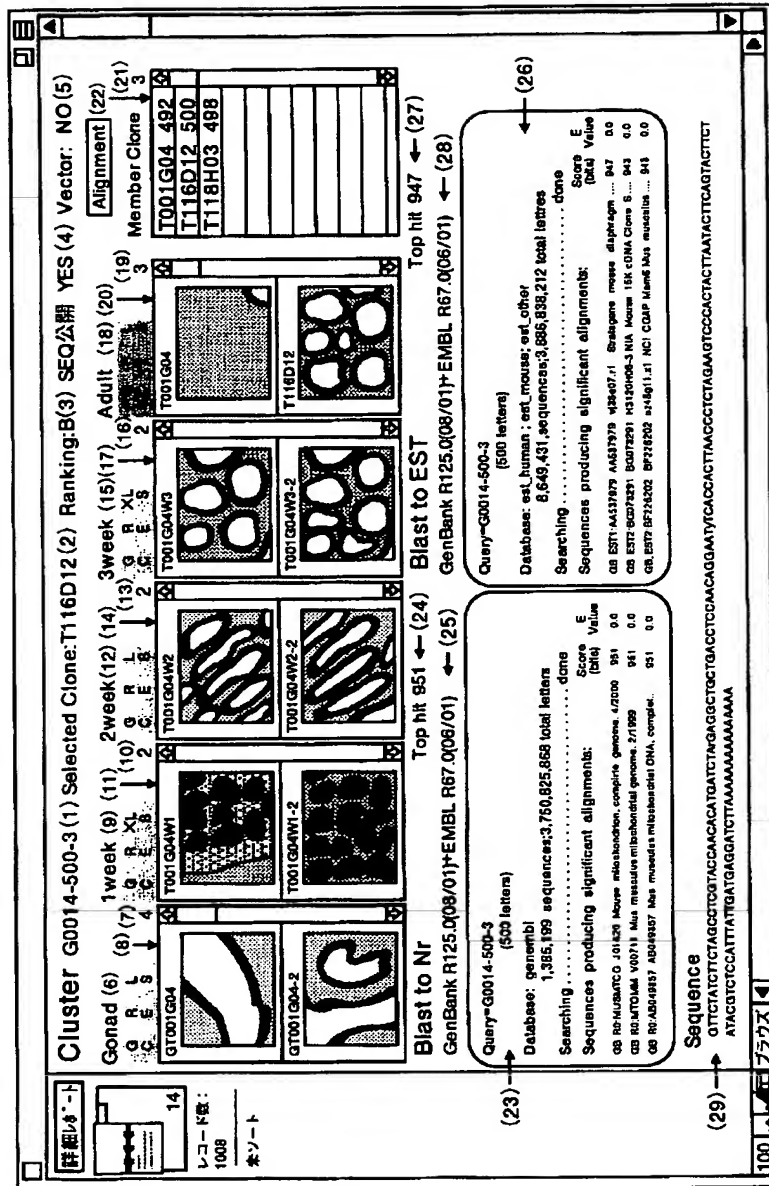
【図17】

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Cluster	Go	1w	2w	3w	Ad	Total	Clones	Selected Clones	Top hit (n)	Top hit (EST)	Ranking	Vector
G0001-711-2	0	1	1	1	2	5	2				C	NO
G0002-840-4	0	0	0	0	3	3	4		220	220	R	NO
G0003-578-2	0	0	0	0	1	1	2		92	101	R	NO
G0004-649-2	5	2	2	2	2	13	2				C	NO
G0005-548-2	0	0	0	0	2	2	2		205	189	R	NO
G0006-631-2	0	0	0	0	3	3	3				R	NO
G0007-581-2	0	1	1	1	2	5	2		107	809	C	NO
G0008-549-21	2	2	2	2	18	26	21	T093G07	1055	1055	B	NO
G0009-585-10	0	1	1	2	9	13	10	T045C04	1189	1068	B	NO
G0010-536-2	0	2	1	1	1	5	2		119	100	C	NO
G0011-514-7	5	2	2	2	7	19	7	T002Q10	984	984	B	NO
G0012-528-18	0	1	2	2	13	18	18	T001A06	1001	997	B	NO

【図19】



【図18】



【手続補正書】

【提出日】平成15年4月2日(2003.4.2)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0325

【補正方法】変更

【補正内容】

【0325】さらに、本発明によれば、塩基配列の存在する染色体上の遺伝子座を特定し、染色体の遺伝子座に塩基配列に関する情報(例えば、画像データ、塩基配列、発現量、採取した組織に関する情報、採取した組織

の発育段階または加齢段階に関する情報、遺伝子の発現が見られるか否かに関する情報、および、発現が見られる部位に関する情報など)をマッピングすることにより染色体マップを作成することができる遺伝子発現情報管

理装置、遺伝子発現情報管理方法、プログラム、および、記録媒体を提供することができる。

【配列表】

SEQUENCE LISTING

```

<;110>; Celestar Lexico-Sciences, Inc.
<;120>; Gene expression information management apparatus,
method, program, and storage medium
<;130>; PCLA-14009
<;160>; 3
<;170>; PatentIn version 3.1
<;210>; 1
<;211>; 157
<;212>; DNA
<;213>; Mus testes
<;220>;
<;221>; Inventor: Yasuo, Uemura; Hirofumi, Doi

<;220>;
<;223>; DNA Sequence 1
<;400>; 1
gtttctatctt ctacgctcgt accaacacat gatctargag gctgctgacc tccaacagga    60
atytaccac ttaaccctct agaagtccta ctacttaata cttcagtact tctatacgtc    120
tccatttatt gatgaggatc ttaaaaaaaaa aaaaaaa    157

<;210>; 2
<;211>; 218
<;212>; DNA
<;213>; Mus musculus
<;220>;
<;221>; misc_feature
<;222>; (167)..(167)
<;223>; n stands for any base
<;220>;
<;221>; misc_feature
<;222>; (197)..(197)
<;223>; n stands for any base

<;220>;
<;221>; misc_feature
<;222>; (202)..(202)
<;223>; n stands for any base
<;220>;
<;223>; DNA Sequence 2
<;400>; 2
atataataga cgaaatcaac aaccccdtat taaccgttaa aaccataggg caccaatgat    60
actgaagcta cgaatatact gactatgaag acctatgctt tgattcatat ataatgtgga    120
tctaaccata gcttatgccc arrgtcctag aaatggttcc acaaaanatt tcgaaaactg    180
arctgcttca ataattnaat tnaaaaaaaaa aaaaaaa    218

<;210>; 3
<;211>; 157

```

(40) 103-242154 (P2003-U京穀)

<;212>; DNA
<;213>; Mus testes
<;220>;
<;223>; DNA Sequence 3
<;400>; 3
gttctatctt ctagcctcgt accaacacat gatctargag gctgctgacc tccaacagga 60

atytaccac ttaacctctt agaagtcca ctacttaata cttcagtact tctatacgtc 120
tccattttatt gatgaggatc ttaaaaaaaa aaaaaaa 157

フロントページの続き

(72)発明者 土居 洋文
千葉県千葉市美浜区中瀬1丁目3番地 幕
張テクノガーデンD17 セレスター・レキ
シコ・サイエンス株式会社内

Fターム(参考) 4B024 AA20 CA04 CA09 HA11 HA19
4B029 AA23
5B075 ND20 UU19